Investigation of Protein Biomarkers in HIV positive and HIV negative associated DLBCL

Lerato Hlatshwayo

The thesis presented for master's degree.

In the division of Anatomical Pathology

University of Cape Town

06 February 2020

Contents

Declaration

DECLARATION

I, **LERATO HLATSHWAYO**, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Acknowledgement

I would like to thank my supervisors Prof. Richard Naidoo and Dr. Dharshnee Chetty for laboratory facilities and the offered opportunity to conduct research at the Division of Anatomical Pathology. They have been really supportive throughout this journey.

I further wish to applaud Subash our lab technician and Daniel our current Ph.D. student for helping me with sample preparation and for their continuous support through my MSc journey. Daniel played a significant role, by investing his time and assisting me this optimizing and mastering the proteomic techniques.

My MSc journey not been a smooth one, I appreciated Dr. Henry for stepping up and pushing me to the finish line. I would have not completed my experiments without his support and guidance. Mrs. Nandi from the hair and skin research lab has always made time train guide me through the MALDI experiments, I truly appreciate that. I would like to thank our collaborators Dr. Nelson and Dr. Bridget for advising us on experimental design and experimental planning.

Lastly, I would like to thank my mother for always showing me love and support throughout my MSc journey. I would not have completed this MSc without her positive energy and guidance.

Abbreviations

Acetonitrile (ACN),

Activated B-cell-like (ABC)

B-cell leukemia/lymphoma-2 gene (BCL-2)

Combination antiretroviral therapy (cART)

α-cyano-4-hydroxycinnamic acid (CHCA)

cell of origin (COO)

Diffused large B cell lymphomas (DLBCLs)

Electrospray ionisation (ESI)

Electron capture dissociation (ECD)

Epstein-Barr virus (EBV)

False discovery rate (FDR)

Formalin fixation paraffin embedding (FFPE)

Filter-aided sample preparation (FASP)

Fluorescence in situ hybridization (FISH)

Fourier transform ion cyclo-tron resonance (FT-ICR)

Forkhead box protein P1 (FOXP1)

Gene expression profile (GEP)

Germinal centre (GC)

Groote Schuur hospital (GSH)

Hematoxylin & Eosin (H&E)

Higher-energy collision dissociation (HCD)

Human immunodeficiency virus (HIV)

Indium tin oxide (ITO)

Inhibitory factor 1(IF1)

International prognostic index (IPI)

Label-free quantitation (LFQ)

liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS)

Mass-to-charge (m/z)

Matrix-assisted laser desorption/ionization (MALDI)

mass spectrometry (MS)

National Cancer Center Network (NCCN)

National Health Laboratory Service (NHLS)

non-Hodgkin lymphoma (NHL)

Principal component analysis (PCA)

Positron Emission Tomography (PET)

Post-translational modifications (PTMs)

Reactive lymph nodes (RLN)

Regions of Interests (ROIs)

Rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP)

Rituximab, Cyclophosphamide, Hydroxydaunorubicin, Oncovin and Prednisolone (R-CHOP)

Room temperature (RT)

Time of flight (TOF)

World Health Organization (WHO)

Abstract

Introduction: Diffused large B-cell lymphomas (DLBCL) is the most common aggressive non-Hodgkin lymphoma (NHL) worldwide, constituting up to 40% of all cases globally. The incidence of HIV-associated lymphoma has decreased since the introduction of combination antiretroviral therapy (cART) in the mid-1990s. However, NHL, especially DLBCL remains the most common cause of morbidity and mortality among people living with HIV/AIDS, especially in sub-Saharan Africa where 70% of the global HIV/AIDS population reside. Gene expression profiles (GEP) identified based on the cell of origin (COO) two distinct DLBCL subtypes; germinal-centre B-cell-like (GCB) and activated B-cell-like (ABC) DLBCL. These subtypes differ in their genetic abnormalities and response to treatment regimens.

Aim: We aimed to investigate in detail, protein distribution profile from FFPE tissue in HIV and non-HIV related DLBCL subtypes.

Methods: FFPE DLBCL lymph node tissue samples from HIV and non-HIV related DLBCL were subjected to MALDI-imaging, in order to get the spatial distribution of proteins in DLBCL tissue. Proteins were extracted from tissue samples and subjected to liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) to identify proteins present in FFPE DLBCL tissue. The protein profiles from the above-mentioned samples were compared and characterized by cancer pathways.

Results: This study had 12 DLBCL cases and 2 human tonsil controls diagnosed from 2009-2011. These cases were retrieved using the NHLS database. The overall age of DLBCL patients by the time they were diagnosed ranged from 18 to 73 years, with a median age of 48 years. MALDI-IMS peak detection function identified 1466 different m/z values from both the HIV negative and HIV positive DLBCL cases. There were only 50 exclusive m/z values that distinguished the DLBCL subtypes, Using LC-MS/MS we identified a total of 88 proteins, by comparing these proteins, we observed 6 differentially expressed among the DLBCL subtypes and controls Fructose-bisphosphate aldolase C was the only significantly differentially expressed proteins between HIV negative ABC DLBCL and HIV positive ABC DLBCL subtype (p value=1,47738).

Conclusion: Using proteomic techniques, we identified and visualized differentially expressed protein in DLBCL subtypes and controls. The majority of these proteins belonged to glycolysis, ATP synthesis, and cellular movement.

1. Introduction

The war against cancer is approaching its 50th anniversary, and yet cancer remains the second leading cause of death globally. According to the World Health Organization (WHO), cancer was liable for an estimated 9,6 million deaths in 2018 [1]. Recent discoveries from the Human Genome Project have improved our understanding of the disease, with the use of immunotherapy acknowledged as the fourth pillar of cancer treatment, along with radiation, surgery and chemotherapy [2].

Cancer is a complex disease, characterised by the proliferation of abnormal cells (known as the tumour). These cells undergo mutations caused by different cancer-causing agents, such as; environmental agents, viral or genetic factors [3]. For example, tobacco is associated with lung, bladder, mouth and throat cancers [4].

Uncontrollable cell growth is the hallmark of cancer[5]. This is mainly due to the activation of oncogenes, which allows the proliferation and metastasis of abnormal cells[5,6] (figure.1).

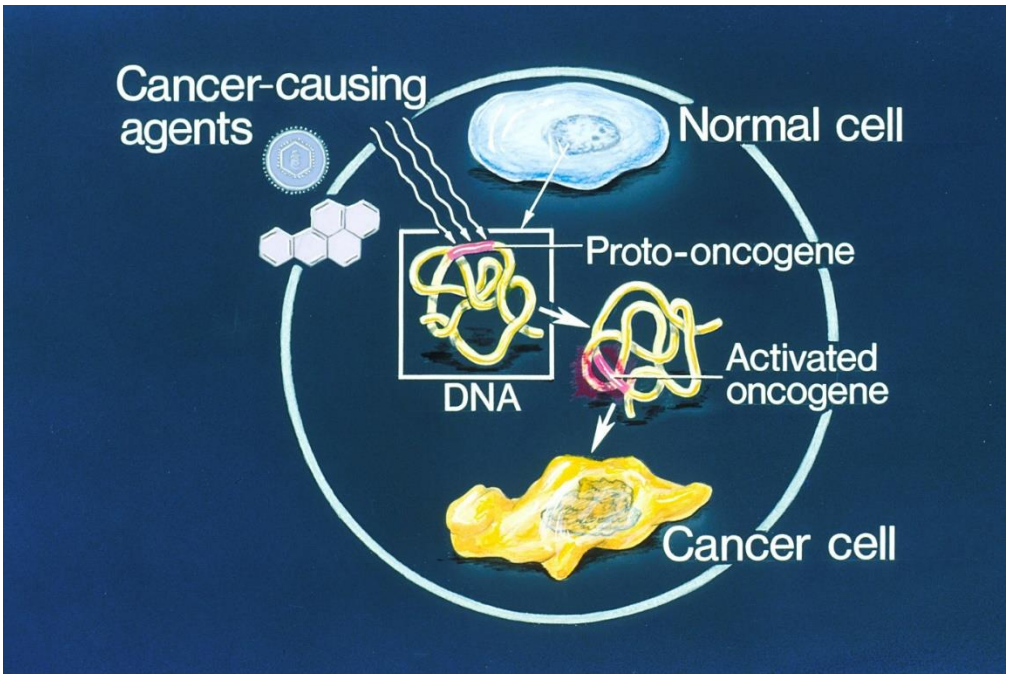


Figure 1: **Oncogene activation,** shows how cancer genes (oncogenes) become activated by cancer causing agents. (Adapted from ref,[5]).

Cancer is usually diagnosed at “advance stages” (usually stages 3 and 4), and treatment is only effective in a few patients. Therefore, the discovery of clinically significant biomarkers can be used for diagnosis, prognosis and potential drug targets [7]. The purpose of this study is to use modern proteomics technology to compare the differences in protein expression profiles of Diffuse Large B-cell Lymphoma (DLBCL) in the hope of identifying new biomarkers.

1.1 Diffuse Large B-cell Lymphoma

Diffuse large B cell lymphoma (DLBCL) is the most common type of non-Hodgkin lymphoma (NHL) and is diagnosed in approximately 150,000 people annually[8]. According to the American Cancer Society, in 2017 there were 80 500 newly diagnosed lymphoma cases and 72 240 of those cases were NHL in the USA, thus ranking NHL among the top causes of death in the USA[9]. DLBCL is prevalent in young HIV positive patients and elderly patients, with the median age range of 60 to 80 years [10].

DLBCL is characterised by the diffuse growth patterns of large and mature B-lymphocytes[11]. These cells are usually twice the size of normal lymphocytes[12]. The disease develops in the B-lymphocytes, which are responsible for the production of immunoglobulin (Ig) in the lymphatic system[12]. Clinically, patients with DLBCL present with lymphadenopathy, caused by rapidly growing cancer cells in the lymph nodes. B symptoms such as weight loss, fever, and night sweats are associated with DLBCL, however, some patients present with symptoms that involve the organs [13–15].

The aetiology of DLBCL in most cases is unknown. However, the pathogenesis of the disease involves the interplay between biological (sporadic or inherited mutations) and environmental (carcinogens and pathogens) factors [16]. Biological and environmental factors can act individually or in sequence to cause genetic mutations [6]. Epstein-Barr virus (EBV) is an example of a cancer-causing virus known to be associated with DLBCL[17]. The presence of EBV during the development of cancer is to induce cell proliferation and lymphangiogenesis through the activation of the NF-κB pathway [18,19]. A study by Herndier and colleagues observed that the presence of both HIV and EBV result in B-cell immortalisation, activation of EBV, and dysregulation of MYC[20]. EBV infection is common in the ABC DLBCL subtypes, and has an impact on the overall survival of patients[21,22].

The use of combination antiretroviral therapy (cART) in the mid-1990s has reduced the incidence of HIV-associated lymphomas in young patients [16,23]. However, NHL, especially DLBCL remains the most common cause of morbidity and mortality among people living with HIV/AIDS, especially in sub-Saharan Africa where 70% of the global HIV/AIDS population resides [24]. Little is known about how HIV induces lymphoma in HIV-associated lymphoma. Multiple factors that may induce lymphoma include oncogenic viruses, immune suppression, expression of cytokines that result in the proliferation of B-cells, or opportunistic infections such as EBV and herpesvirus[22,2512,26,27]. Various studies suggest that low CD4 count increases

the presence of oncogenic viruses thus inducing abnormal cell growth [26,27]. HIV-associated DLBCL has a high expression of immunoglobulin genes (Ig) and this results in the expansion of polyclonal and monoclonal B-cells [28]. A study by Landgren found a high expression of serum immunoglobulin light chain in HIV-associated lymphomas [29]. Thus, suggesting the use of serum immunoglobulin light chain as a prognostic marker for HIV-associated DLBCL[29].

1.2 Classification

The WHO classification system in 2008 grouped DLBCL cases based on clinical and pathological features, into EBV positive DLBCL of the elderly, primary DLBCL of the central nervous system, T-cell/histiocyte-rich large cell lymphoma, and primary cutaneous DLBCL, leg type. However, the 2016 WHO classification system was revised and the DLBCL cases were classified based on the cell of origin, MYC, and BCL3 expression, as well as the EBV status[30].

1.2.1 Cell of origin (COO)

On the bases of gene expression profiling (GEP), DLBCL is classified based on the COO into the germinal centre (GCB DLBCL) and activated B-cell (ABC DLBCL) subtype (figure 2) [31]. The COO reflects the stage at which the disease developes or the biological pathways involved in the development of the disease. The GCB subtype expresses the germinal centre associated genes, such as CD10, LM02, and BCL6 [32–34]**.** The genetic alterations associated with GCB DLBCL cases include *c-rel (*critical regulator) amplification, chromosomal translocation t (14; 18), and histone methyltransferase EZH2 mutations [24–26]. Somatic histone mutations within the EZH2 genes can work together with BCL6 genes to initiate the growth and development of GCB-DLBCL[39,40]. Various studies suggest the use of both EZH2 and BCL6 genes as selective drug targets for GCB DLBCL[40–42]. The cellular energy metabolism and growth of DLBCL are regulated by the activation of phosphatidylinositol 3 kinases due to PTEN gene mutation in both GCB DLBCL and ABC DLBCL [38,39]. The expression mechanism of BCL2 differs in both the GCB DLBCL and ABC DLBCL. In the GCB subtype, BCL2 expression occurs through chromosomal translocation t (14; 18), while it occurs through gene amplification and transcriptional modification in the ABC subtype [39].

The ABC DLBCL subtype expresses plasmablastic cells like genes such as FOXP1, and MUM1. The hallmark of the ABC DLBCL subtype is the activation of the NF-kB pathway and amplification of SPIB and INK4a/ARF genes [26,38,42]. The NF-kB pathways are activated by the

CARD11–BCL10–MALT1 (CBM) signalosome complex. The CBM complex links and triggers the activation of B-cell receptors with the NF-kB pathway[44]. The CD79A gene mutation and Bruton tyrosine kinase regulate the B-cell receptors signalling pathway [38,43]. The upregulation of Bruton tyrosine kinase is associated with MYD88 mutation[41,45].

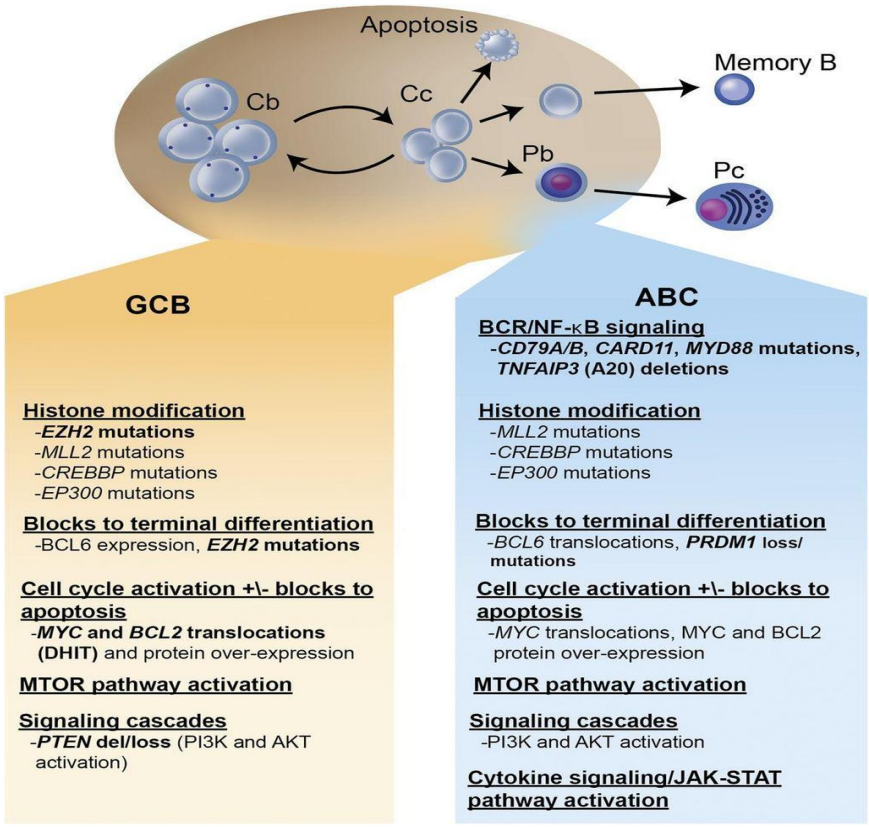


Figure 2**. DLBCL subtypes based on the cell of origin.** GCB and ABC subtype arise at different stages of B-cell development in the germinal centre. GCB arises from Centroblastic cells, whereas the ABC arises from Plasmablastic cells. Both these subtypes differ in the mutation, oncogenic pathways, and translocation. (Adapted from the ref, [39]).

1.2.2 MYC and BCL2 expression

The revised WHO classification highlight the prognostic significance of double expressors, which is defined as the simultaneous expression of MYC and BCL2 proteins. The suggested threshold for MYC protein expression is more than 40% positive tumour cells, and the threshold for BCL2 protein expression is more than 50% positive tumour cells[30,46]. MYC and BCL2 double expressors have been reported in approximately 34% of DLBCL patients. The double expressors have a worse prognosis than patients that express only one protein. Double expresser cases are more common in the ABC subtypes and this may contribute to the overall survival of ABC DLBCL patients[47].

1.2.3 EBV positive DLBCL

EBV positive DLBCL is defined as a new disease entity[30]. The term "EBV positive DLBCL of the elderly" was used to describe EBV positive DLBCL occurring in elderly HIV negative patients. However, recent studies revealed that EBV positive DLBCL does also occur in younger patients. Hence the term "elderly" is substituted with "not otherwise specified"[48]. There are controversies about the impact of R-CHOP on the prognosis of EBV positive DLBCL patients. Most studies show that EBV negative DLBCL patients respond better to R-CHOP than EBV-positive DLBCL patients[18,22,49].

1.2.4 Burkitt-like lymphoma with 11q aberrations

Burkitt-like lymphoma with 11q aberration are rare cases characterised by Burkitt lymphoma features, however, they lack MYC rearrangement. Instead, they carry the 11q chromosomal alteration with proximal gains and telomeric losses[50,51]. Similar to Burkitt lymphoma, Burkitt-like lymphoma with 11q aberration have aggressive clinical features[50].

1.3 Diagnosis and staging

DLBCL is usually diagnosed from a lymph node biopsy, or excision, based upon clinical examination and diagnostic imaging [52]. A pathologist examines tissue for B-lymphocyte antigens such as CD19, CD20, CD22, CD79, and CD45 [53,54]. Further immunohistochemical (IHC) stains performed CD10, BCL6, and MUM1 in order to classify the cases according to the revised WHO 2016 classification of DLBCL molecular subtypes, using the Hans algorithm[30,46]. DLBCL cases are classified as GCB subtype if CD 10 positive, and as ABC subtype if both CD10 and BCL6 negative. The expression of MUM1 is used to determine subtype if a case is CD10 negative and BCL6 positive. If the case is MUM 1 positive, it is classified as ABC subtypes and classified as GCB subtype if MUM1 negative (figure 3)[55]. A threshold of more than 30% is used to interpret stains as either positive or negative.

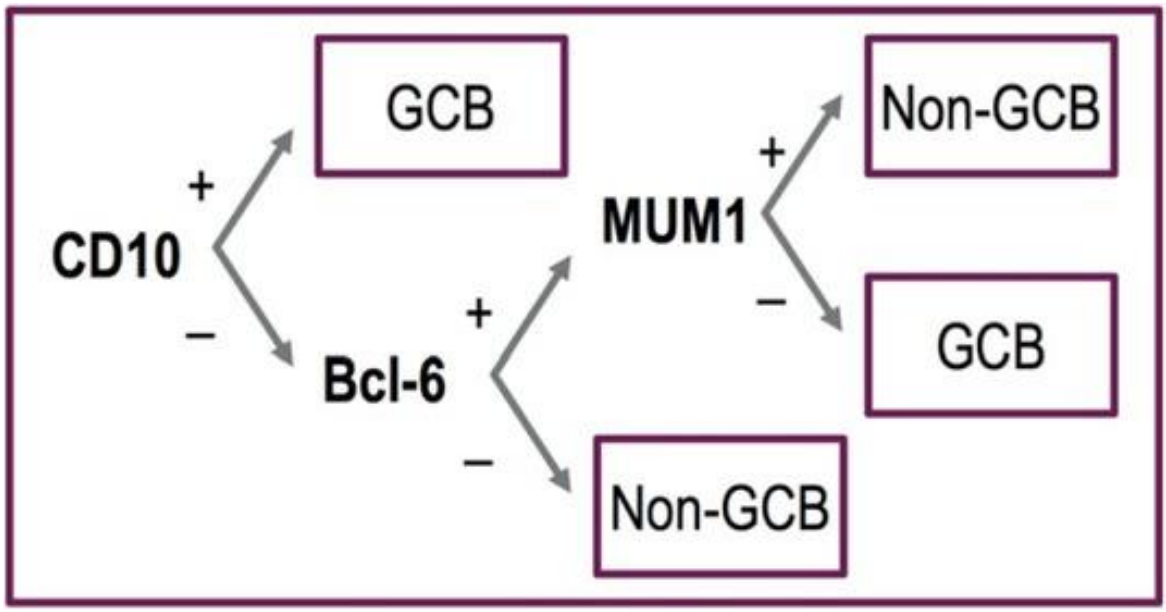


Figure 3. **Hans classification** of DLBCL subtypes according to the immunohistochemical stains. GCB-germinal centre* (Adapted from ref [55])

Tissue examination also includes cytogenetic analysis, molecular genetic analysis and immunophenotypic analysis (Figure 4) [52,54]. The cytogenic analysis (also known as karyotyping) is used for classifying and diagnose cancers based on chromosomal abnormalities[53]. Fluorescence in situ hybridization (FISH) is a commonly used method for cytogenetic analysis and is used to detect MYC, BCL2, and BCL6 gene rearrangements in DLBCL[56]. If tissue expresses both MYC and BCL2 or BCL6, it is diagnosed as high-grade B-

cell lymphoma (previously known as double-hit DLBCL). Triple hit (THL) lymphomas express all three genes (MYC, BCL2, and BCL6)[56]. A molecular diagnostic assay assesses the types of genetic alterations and the total number of alterations present in the tissue. The immunophenotypic analysis assesses the expression of DLBCL diagnostic markers such as MYC, BCL2, MUM1, Ki67, etc [39,52].

Cancer staging is a system that describes cancer based on its invasiveness and metastasis [57,38]. The Positron Emission Tomography (PET) scans used to identify the size and site of the tumour[38]. There are four cancer stages, the "early stage" which includes stages 1 and 2 and the "advance stage" which includes stages 3 and 4 [32,45].

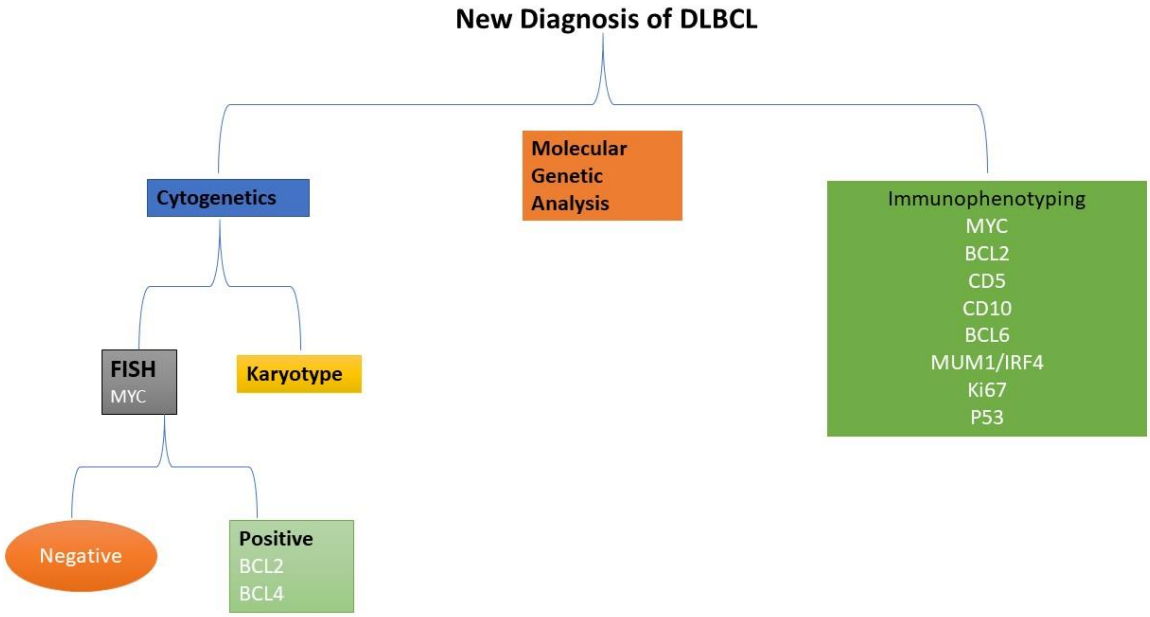


Figure 4. **Diagnostic examinations of DLCBL.** Cytogenetic analysis, molecular genetic analysis and immunophenotypic analysis (adapted from ref,[38])

1.4 Treatment.

DLBCL is usually diagnosed at "advantages stages", therefore, the standard treatment regimen is rituximab, cyclophosphamide, hydroxydaunorubicin, oncovin, and prednisolone (R-CHOP)[58,59]. Although this regimen is given in twenty one day cycles for an average of six cycles, the length of treatment differs with individual patient [60]. Approximately, 60-70% of DLBCL patients respond well to R-CHOP treatment, while the remainder relapse and die [59]. Therefore, there is a need to develop effective salvage therapy for patients that relapse. The R-CHOEP regimen (a combination of etoposide with R-CHOP) in young HIV positive patients have been effective [60,61]. Current treatment strategies aim at developing novel drug agents with reduced drug toxicity and dosage for patients who cannot be cured using R-CHOP[62]. Recently, ibrutinib, a novel drug has been developed to inhibit Bruton's tyrosine kinase, a crucial enzyme for the development of B-cell cancers [41].

1.5 International prognostic index (IPI)

The IPI is a classification system was developed to predict the prognosis of patients with NHLs[63]. This system uses five clinical parameters such as age, Ann Arbor staging, serum lactate dehydrogenase (LDH), oncology performance status, and extra-nodal involvement[63]. The prognosis scoring system has four risk groups. The low risk group has (0 to 1 point), low-intermediate risk group has (2 points), high-intermediate risk group has (3 points), and high risk group has (4 to 5 points)[63]. The National Comprehensive Cancer Network (NCCN) proposed a better IPI system during the R-CHOP era. The NCCN-IPI uses a maximum scoring of eight and redefined the age, LDH, and the extranodal involvement include sites such as bone marrow, lungs, brain, liver, and gastrointestinal tract [64].

1.6 Biomarker

Biomarkers are biological molecules (such as protein, nucleic acids, specific cells or metabolite) that can be used to measure the of a disease severity [65]. These molecules can be used as diagnostic, or prognostic markers for a disease [66]. The discovery of protein biomarkers has increased over the years because proteins are the centre of almost all biological and cellular processes[66]. There are several protein biomarkers that frequently appear in literature and play a role in the clinical diagnosis and prognosis of DLBCL.

B-cell lymphoma-2 *(*BCL2)** is an antiapoptotic protein that regulates cell death[67]. Mutations in the BCL2 gene have been identified in almost all cancers. The overexpression of BCL2 is regulated by MYC[68]. BCL2 is upregulated in the GCB subtype compared to the ABC subtype [69]. The R-CHOP regimen has altered the prognostic significance of previously studied DLBCL biomarkers[70,71]. Therefore, there are controversies about the effect of BCL2 on DLBCL patient outcomes. Various studies suggest that the upregulation of BCL2 is associated with poor patient outcome in the ABC subtypes [67,72,73].

Ki-67 (also known as MKI67) is a nuclear protein, clinically used to measure tumour proliferation of NHL[74]. Proliferation is measured by dividing all Ki-67 positive cells in a sample by the total number of cells present in a sample[75]. High Ki-67 expression is associated with NHL, however, in DLBCL, Ki-67 expression is regulated by MYC and TP53 proteins [73,74,76–78].

B-cell lymphoma 6 gene rearrangements (BCL6) is a common chromosomal abnormality in DLBCL that occurs at chromosome band 3q27. BCL6 rearrangement occurs in at least 50% of DLBCL patients [56]. BCL6 express genes that regulate the development of B-lymphocytes in the germinal centre [56]. The expression of BCL6 is upregulated in the GCB subtype compared to ABC subtype[39]. The presence of both BCL6 and MYC rearrangements are associated with HIV- associated double hit lymphoma [79]

Tumour proteins p53 genetic mutations. The human P53 gene encodes for intracellular tumour p53 proteins[72]. The tumour p53 proteins are found in the cell nucleus and function to repair damaged DNA, by activating DNA repair proteins, as well as tumour suppressor proteins that prevent uncontrollable cell growth [80,81]. The mutation in tumour p53 is found in approximately 20% of DLBCL cases. These mutations are detected by DNA sequencing and they lack CD19 markers [80,82–84]. The expression of tumour p53 proteins is associated with MYC rearrangement and has a poor prognosis of DLBCL[85]. Studies suggest the use of both MYC and tumour p53 proteins as markers for double hit lymphoma[82,86].

MYC rearrangements. MYC genes encode for transcription factor proteins located on chromosome band 8p24 [42,68]. DLBCL cases with MYC translocations and immunoglobulin gene loci have a poor prognosis compared to non-immunoglobulin gene DLBCL cases [42,87,88]. Double hit lymphomas have both BCL2 and MYC gene rearrangements, and the overexpression of these proteins has a poor prognosis [34,68]. In addition to MYC rearrangement, DLBCL cases

may have extra copies of MYC amplification, which have a negative impact on patient survival [42,68].

1.7 Proteomics

Proteomics is the study of the identification and characterization of the entire set of proteins synthesised by the cell [89]. Proteins are the centre of action for almost all cellular processes because they determine the structure and function of cells[89]. Proteins can undergo structural and functional changes during cellular processes, disease conditions, cellular stress or in response to medication. The human genome sequence has made it possible to discover and characterize proteins that are of diagnostic and prognostic value[88]. The role of proteomics in cancer research is to identify protein markers that can be used for early detection of specific cancers; prostate-specific antigen is considered as a good biomarker of prostate cancer [90]. The application of proteomic technology in cancer has led to the discovery of numerous biomarkers for many cancers. However, these biomarkers are not being used in clinical settings because they have low sensitivity and specificity[91].

Proteomic methods are labour intensive and require sample preparation, protein digestion, and identification [92]. Proteomic analysis studies are hindered by biological sample complexity [89]. A lot of fractionation methods have been developed to reduce sample complexity and increase the yield of extracted proteins in FFPE tissue specimen [89,93]. The most commonly used fractionation methods are liquid chromatography (LC) and gel electrophoresis [89,92,94]. The LC approach reduces sample complexity prior to mass-spec analysis by enzymatically digesting proteins to peptides [81]. The peptide eluents are then sequenced with tandem mass spectrometry (MS/MS)[90,95]. The major advantage of the LC approach is that the separation and identification of proteins are faster and simpler than 2DE electrophoresis. The 2DE gel electrophoresis is commonly used in quantitative proteomic studies, and it separates proteins based on the molecular weight or isoelectric point through a polyacrylamide gel [89]. The separated protein fragments are then visualized on the gel using fluorescent dyes and extracted prior to detection on mass-spec analysis [66,89]. The advantage of using a 2DE gel based proteomic approach is its ability to identify intact and hydrophilic proteins [89,96].

1.7.1 Mass spectrometry in Proteomics

Mass spectrometry is a powerful analytical tool used for the quantification and identification of known or unknown compounds in a sample[97]. The mass spectrometer uses an ion source to ionise proteins into gaseous ions. Generated ions are then separated and detected based on their mass to charge ratio (m/z)[88]. Mass spectrometric technology differs in the type of ion source and mass analysers. The commonly used ion sources are matrix assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI) [91,97]. There are four types of mass analysers, these include; time of flight (TOF), ion trap, quadrupole, and Fourier transform ion cyclotron resonance (FTICR)[98]. MALDI is a soft ionization technique and is usually combined with TOF analysers, whereas ESI is normally combined with ion trap analyser [99]. During MALDI experiments a matrix solution, (which is an acidic molecule that absorbs energy from the laser) is applied to tissue sections and it co-crystalizes with the analyte prior to ionization with laser beams [100,101]. The mass spectrum is measured based on the time it takes an activated ion molecule to reach the ion detector. The small ions reach the detector first and the larger ion reaches the detector last (figure 5) [97,102].

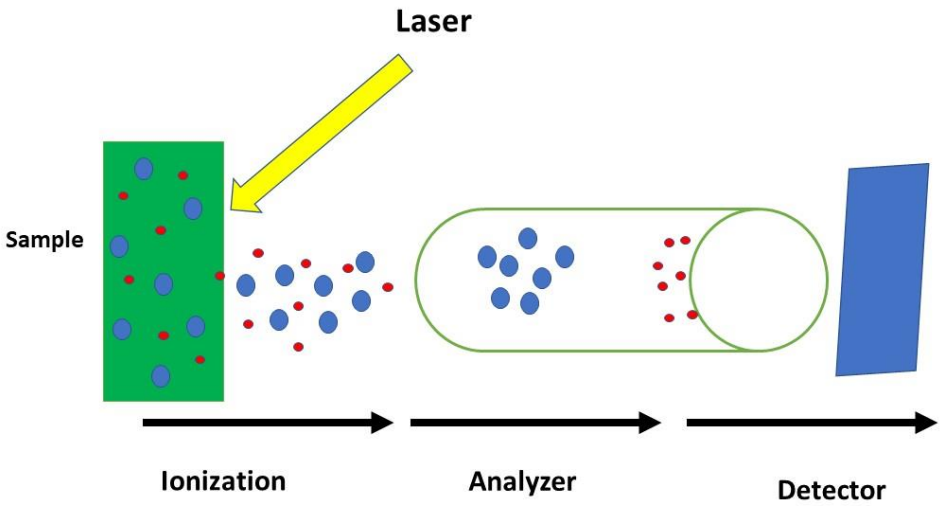


Figure 5. **MALDI processing steps**. The laser absorbing matrix is applied to samples and it forms co-crystals with the sample. Laser beams are subjected to co-crystal analyte samples, which results in ionization and desorption of sample analyte. Finally, ions move to the ion detector where the m/z ratio and intensities are determined. (Modified from ref,[100]).

1.7.2 MALDI-IMS of biological samples

MALDI imaging mass spectrometry (IMS) is a tool used to investigate the spatial distribution of molecules within a tissue section[103,104]. The visual distribution of molecules such as metabolites, xenobiotic, lipids, peptides, and proteins can be determined using this technology[105]. The MALDI-IMS methods differ based on the type of matrix used during experiments, and the choice between matrixes depends on the type of analyte to be studied [98]. Sinnapinic acid is a commonly used matrix to visualise the distribution of intact protein, while α-cyano-4-hydroxycinnamic acid (CHCA) matrix is commonly used to study peptides [106]. The Matrix application can be done manually or automatically using a sprayer or spotter[102]. The HTX TM-Sprayer is a commonly used matrix spraying instrument and it provides a thin homogenous matrix layer[102].

MALDI-IMS technology can be applied on formalin fixed paraffin embedded tissue sections (FFPE), this approach maintains the histological quality of tissues and allows the detection of molecules with a high molecular weight such as intact proteins [107]. However, formalin fixation of tissue causes protein-protein crosslinking due to methylation of amino acid side chains [70]. Sever formalin treatment affects the extraction of soluble proteins and also causes DNA fragmentation, and this can hinder proteomic and genomic data analysis [109,110]. Despite this, various MALDI-IMS protocols that aim to reverse the cross-linking proteins have been developed [107,111]. Their workflow involves the use of citrate or Tris-EDTA heat-induced antigen retrieval and enzymatic digestion of proteins present in FFPE tissue samples. These protocols have been effective in reducing sample complexity [112]. Despite this, FFPE tissue specimen are a valuable resource to translational proteomic studies because the samples have a well-documented clinical data[107]. Although cancer research publications have taken advantage of MALDI-IMS technology, the use of this technology in a clinical setting is not cost effective [100,113].

2. Aims and Objectives

The overall aim of the study is to investigate in detail, the protein distribution profile and assess these profiles in HIV and non-HIV related DLBCL FFPE tissue using MALDI- Imaging and LC-MS/MS analysis. The specific objectives were to:

1. Prepare FFPE DLBCL tissue for proteomic analysis,
2. determine the spatial distribution of protein signatures in tissue sections using MALDI IMS,
3. identify proteins present in FFPE DLBCL tissue using liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS),
4. compare the protein profiles isolated from the abovementioned specimens, and
5. investigate differences in biological pathways of the proteins across cohort.

3. Materials and Methods

3.1. Human Research Ethics Approval

This study was approved by the Human Research Ethics committee of the Faculty of Health Science, University of Cape Town (HREC Ref: 568 /2018).

3.2 Study design

Based on previously published data [91,114], the present retrospective study consisted of 12 archived FFPE DLBCL tissue samples [DLBCL from HIV negative patients (n=6) and DLBCL from HIV positive patients (n=6)]. The archived FFPE DLBCL tissue samples were further classified into the germinal centre B-cell-like (GCB) and activated B-cell-like (ABC) subtypes, according to the Hans Choi algorithm (figure 6). Human tonsil samples were used as controls for both HIV positive and HIV negative patients. The HIV positive human tonsil was tested for p24 to confirm the HIV status, while the HIV negative human tonsil was added to compare any shared or common biomarkers present.

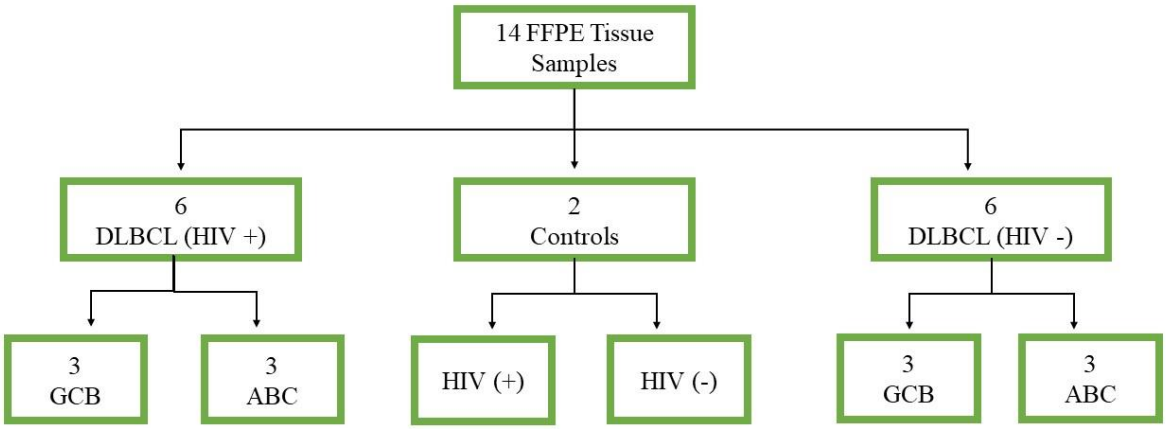


Figure 6. **Summary of the Study design.** Study consisted of a total of 14 archived FFPE sample that were grouped as 6 DLBCL [HIV negative (n=3ABC; n=3 GCB)], 6 DLBCL [HIV positive (n=3ABC; n=3 GCB)] and 2 human tonsil controls (n=1 HIV positive); n=1 HIV negative).

3.3 Case selection and Data collection

The DISA database in the Division of Anatomical Pathology/ National Health Laboratory Service (NHLS) was used to search for DLBCL cases. The study cases were confirmed by a Pathologist by re-examining the H&E histology slides. All EBV positive cases were excluded. Cases with plasmablastic lymphomas, Burkitt's lymphoma and B-cell lymphomas with intermediate features between large B-cell lymphoma and Burkitt's lymphoma were aslo excluded.

3.4 Tissue preparation

Tissue was prepared differently for MALDI-IMS and LC-MS/MS.

3.4.1 Tissue preparation for MALDI-IMS

Sections of 10-micron thickness were cut on a rotary microtome (LEICA RN 2125 RTS) from FFPE tissue blocks and transferred onto indium tin oxide coated (ITO glass) slides (Sigma-Aldrich). The slides were then heat fixed on a hot plate at 60 °C for 10 mins. Sections were dewaxed in three jars of xylene (Merck, SAAR221120LC) for 5 min each. They were then cleared through decreasing concentrations of ethanol (Kimix chemical and lab supplies, AC003) and washed well in water. This was followed by the antigen retrieval process, which was incorporated into our method to reverse the cross-linking of protein due to formalin fixing.

3.4.2 Tris-EDTA Buffer Antigen Retrieval

A pressure cooker containing Tris-EDTA [2.42g tris base (Sigma, BCBS3963V) and 0.74g EDTA (Sigma, SLBH4839V)] buffer solution (pH 9) was preheated until the temperature reached 95-100 °C. Slides were then immersed in the pressure cooker and incubated for 90 sec. They were then removed, washed well in water and allowed to dry prior to trypsinization.

3.4.3 Sample tissue digestion and matrix application

A trypsin solution containing (5 ng/µl in 50% acetonitrile; 50 mM ammonium bicarbonate, pH 8) was automatically sprayed (HTX TM-sprayer (Bruker) onto the tissue sections, according to the optimized in-house protocol with a flow rate of 0.0075 ml/min. Twelve coatings were applied onto the tissue. Spraying was carried out for 15 seconds and the drying time was 30 seconds between each coating. The slides with digested tissue were incubated for 4hrs at 37 °C.

Following digestion, slides were automatically coated with α-cyano-4-hydroxycinnamic acid (HCCA) matrix solution containing (7g/L HCCA in 50% acetonitrile, 0.2% trifluoroacetic acid) at a flow rate of 0.15ml/min. The matrix application was done using an optimized HCCA setting on the TM-Sprayer system (HTX Technologies, LLC, Carrboro, NC, USA). The Bruker peptide calibration standard II was used for external calibration. A slide scanner (MF500 Reflecta High-Resolution Tissue Scanner) was used to scan the tissue, and to ensure proper positioning of slides on the Bruker MALDI plate adaptor (MTP Slide Adaptor II).

3.4.4 MALDI-IMS

MALDI-IMS was performed on a RapifleX MALDI TOF/TOF tissue-typer (Bruker) that had flexControl and fleximaging 3.0 software (Bruker Daltonik) installed within the instrument to control the laser beam. The laser beams were rastered onto the tissues slides at 200 × 200 μm2 pixel sizes. The instrument was run at positive ion reflector mode and the detection of spectral peaks ranged at m/z 600–3,500. The reflector voltage was set at 26.71 kV, with a suppression of 400 Da. The acceleration voltage, lens voltage, and final acceleration voltage were set at 25–22.45 kV, 8.00 kV, and 13.35 kV respectively.

3.4.5 MALDI-IMS Data Acquisition

Fleximaging v.3.0 Software (Bruker) is a software used for visualization of the spatial distribution of biomarkers in a tissue sample, by the automatic spectral acquisition and reassembling of ion images[101]. The spectral acquisition parameters were set for each spatial coordinate and these parameters were 2,500 shots per pixel area, with 200-shot increments at a laser frequency of 200 Hz. FlexAnalysis v.3.0 software (Bruker Daltonik) is a Bruker data analysis software that was used for spectral processing and baseline analysis. SCiLs Lab was used for statistical analyses on pooled spectral data from each sample group.

3.4.6 SCiLs Lab data analysis.

SCiLs Lab software (SCiLs Lab 2015b) was used for unsupervised characterization of m/z values in DLBCL subtypes and controls[115,116]. MALDI-IMS raw data was first imported and converted into SCiLS H5 format. The following pre-processing data parameters were performed; normalization, performed at total ion count (TIC), baseline correction, done at iterative convolution with (sigma 20 and 15 iterations), spectra smoothing and spatial segmentation, by grouping similar spectra using clustering algorithms. The Orthogonal Matching Pursuit algorithm was used to select peaks. This algorithm selected 15 peaks per

spectrum and took the consensus peaks that were in at least 1% of the considered spectra[117]. The median filtering and Chambolle algorithm (with lambda 0.5) were used to carry out image denoising. The denoised data were then spatially segmented using dissecting k means algorithm.

3.4.7 Tissue preparation for LC-Ms/Ms

FFPE tissue sections were cut at 10 microns, transferred onto coated slides (Histobond-Marienfeld Lasec) and heat fixed on a hot plate at 60 °C for 10 mins. Sections were dewaxed and antigen retrieved as mentioned above. Once left to dry at room temperature the tissue sections were scraped using scalpel blades and transferred into Eppendorf tubes.

3.4.8 The filter-aided sample preparation (FASP)

The FASP was performed according to previously published protocols with minor medication[91,108,118]. Briefly, all buffer washes were carried out by centrifugation at 14000g for 15 min. Proteins were extracted (50 µL) with three rounds of 200 µL Urea (UA) buffer [8 M UA (Sigma, U5128) in 0.1 M Tris, pH 8.5]. To reduce the alkylation of cysteine bonds on extracted proteins, we incubated samples in 100 µL of iodacetamide (IAA) buffer containing IAA 0.05 M in UA (Sigma) for 20 min in the dark. To remove the alkylating agents, the samples were washed with two rounds of 100 µL UA, followed by three rounds of 100 µL of ammonium bicarbonate buffer (ABC buffer) (0.05M, pH 8). ABC buffer (40 µL) was added to trypsin (Promega) at a ratio of 1:100. Proteolysis was carried out at 37 °C for 18 hrs in a wet chamber.

3.4.9 Desalting

After proteolysis, peptides rich solution was eluted thrice with 50 µL of ABC buffer solution. Desalting was done using a homemade stage tip containing Empore Octadecyl C18 solid-phase extraction disk (Supelco)[119]. The activation, equilibration, peptide wash, and elution of C18 disk were done by centrifugation at 4000 rpm for 1 min. The activation and equilibration of C18 disk were carried by three rinses with 80% acetonitrile (ACN), followed by three rinses with 2% ACN, respectively. To desalt the peptide, 10µL of the peptide-rich solution was added to the C18 disk, followed by three washes of 2% ACN containing 0.1% formic acid (Sigma). Desalted peptides were eluted into a glass insert using three rounds of 50 µL of 60% ACN, 0.1% formic acid. The peptides were then vacuum dried in a SpeedVac vacuum concentrator (Thermo Fisher Scientific ^{TM}) and stored at - 20 °C freezer until ready to be measured by LC–MS/MS.

3.4.10 LC−MS/MS

Vacuum dried peptide in a glass insert were resuspended in 250 ng/µL of 2% ACN, containing 0.1% formic acid, sonicated for 3 min and transferred to a glass autosampler vial prior to mass-spectrometer analysis. 10 µL from each sample were analysed on Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass Spectrometer (Thermo Fisher) with a flow rate of 250 nl/min. The M/S was done in a data-dependent manner and the automatic scan was achieved by switching each M/S scan with 10 MS/MS at a scan range of at 300–1650 *m/z* and with a maximum injection time of 30 seconds. The fragmentation of ion was done at high-energy collision dissociation with the collision energy set at 25 NCE. The mass spectra were acquired at a resolution of 70 000 at a maximum integration time of 250 ms. Ion selection was done at an intensity threshold of 0.001 with charge exclusion of z = 1 ions.

3.4.11 Maxquant LC-MS/MS data processing

Maxquant (v.1.3.0.5) a quantitative proteomic software was used to process all MS raw data acquired from LC-MS/MS runs[120]. Peptide identification was done using the andromeda search engine, which uses a human proteome to search peptides against the International Protein Index (IPI human version 3.87). Trypsin and Carbamidomethylation were selected as enzyme specification and as a fixed modification during the peptide search. The false discovery rate (FDR) for both peptides and proteins was set at 1%. Label-free quantitation (LFQ) on at least one peptide per protein was performed using the MaxLFQ algorithm and the re-quantify function. Minimum cut off for peptide length was set at seven amino acids, and maximum permissible missed cleavage was set at 2. All the LFQ data obtained from Maxquant was then imported into the Perseus software (version 1.6.10.43) for further statistical analysis and visualization by hierarchical clustering.

3.4.12 Perseus: Data processing and Normalization

Perseus is a proteomic software used for the statistical analysis of "omics" data and for the interpretation of protein quantification[121]. The Maxquant output file was first converted into a text (ProteinGroup*.txt) format and loaded to Perseus software (version 1.6.10.43). The LQF intensities from all the samples were loaded as the main columns. The data processing parameters were used to filter, transform and group sample proteins. Proteins that were only identified by site, reverse protein, and potential protein contaminants were filtered out. Samples were grouped according to their replicates and the transformation of data was done using the formula "log2(x)". Data transformation was done in order to normalize the data prior to

applications of the statistical test. A two-sample t-test (-log transformation P-values) set at a threshold of 0.05 (5% probability error) was performed to determine which proteins were significantly different among the DLBCL subtypes. The differential expression levels were visualised by hierarchical clustering. Appendix H, figure 26 summarises the data processing pipeline.

3.4.13 Haematoxylin & Eosin (H&E) Staining for the two tonsil controls.

Sections of 3 microns were prepared from archived FFPE tissue blocks and transferred onto normal microscopic slides (Lasec). After being heat fixed on a hot plate at 60 °C for 10 mins sections were dewaxed through three jars of xylene (Merck, SAAR221120LC) for 5 min each. They were then passed through decreasing concentrations of ethanol (Kimix chemical and lab supplies, AC003) (100, 95, and 70%) followed by washing well in water. Slides were then stained with Haematoxylin (Mayers) for 5 min and blued in ammoniated water. After washing well in water, the slides were placed in Eosin (MERCK) for 2 minutes. Finally, the slides were dehydrated, cleared and mounted on coverslips (Lasec) with entellan (MERCK).

4. Results

4.1 Clinical and biological parameters

This study had 12 DLBCL cases and 2 human tonsil controls diagnosed from 2009-2011. These cases were retrieved using the NHLS database, however, only 11 patient folders were located from Groote Schuur hospital (GSH) (table 1). The overall age of DLBCL patients by the time they were diagnosed ranged from 18 to 73 years, with a median age of 48 years. The number of HIV positive males was 5 and their age at the time of diagnosis ranged from 18 to 54 years, with a median age of 40 years. There were five HIV negative females and their age by the time of diagnosis ranged from 37 to 73 years, with a median age of 55 years. All the cases presented with extranodal lymphomas. The common extranodal sites in both HIV positive and HIV negative cases were submandibular, axilla, mesentery, supraclavicular, testis, and inguinal (Appendix I, table 9). We only included EBV negative cases in this study, because EBV positive DLBCL is another disease entity of lymphoma.

Table 1. The clinical and biological parameters of DLBCL from HIV positive and HIV negative patient study cohort.

		Case (n=12)	**Control (n=2)**
Age range (median) 18 to 73 (48)	**HIV (-) Age range (median)**	Range (37 to 73) Median (55)	**Unknown**
	HIV (+) Age range (median)	Range (18 to 54) Median (40)	
Gender	**M**	6 (5 HIV(+):1 HIV(-))	-
	F	5 all HIV(-)	-
	Unknown	1	-
Classes (HIV Status)	**ABC (Pos: Neg)**	6(3:3)	N/A (1:1)
	GCB (Pos: Neg)	6(3:3)	
EBV status	**Positive**	0	0
	Negative	12	2
Site	**Nodal**	0	0
	Extra-nodal	11	2

The patient folder for one case was not found. Hence the gender, age, and lymph node site were not recorded.

4.2 MALDI IMS results: Principle Component Analysis of Mass-spectral peaks between DLBCL cases and controls were distinguished by MALDI-IMS.

4.2.1 Principle Component Analysis (PCA) Clustering between GCB (HIV+) and HIV (+) control

The FFPE tissue region of interests (ROIs) from HIV (+) GCB DLBCL subtype and HIV (+) control were marked by the pathologist on H&E slides and were superimposed on scanned ITO slides (Appendix C.1, figure 17and 18). SCiLs lab software was used to superimpose the marked ROIs on scanned ITO slides (figure 7A). The average spectra of the ions were observed, and the peak picking function identified 457 ions in the range of m/z 606-3500 (figure 7B). The PCA plots were used to differentiate extracted MALDI mass spectra from three HIV (+) GCB DLBCL samples and one HIV (+) control. The PCA score plot indicated four distinct clusters between the ion patterns of the mass spectra of the three GCB DLBCL HIV (+) samples and one HIV (+) control. The HIV(+) control (indicated in yellow) did not cluster with any of the HIV (+) GCB DLBCL samples (figure 7C). The loading plot of the PCA analysis provided information regarding the contribution of each ion signal to the variance covered by each principal component PCA1 and PCA2. There were 21 different ion signals that contributed to the variance covered by PCA1 and PCA2 (figure 7D).

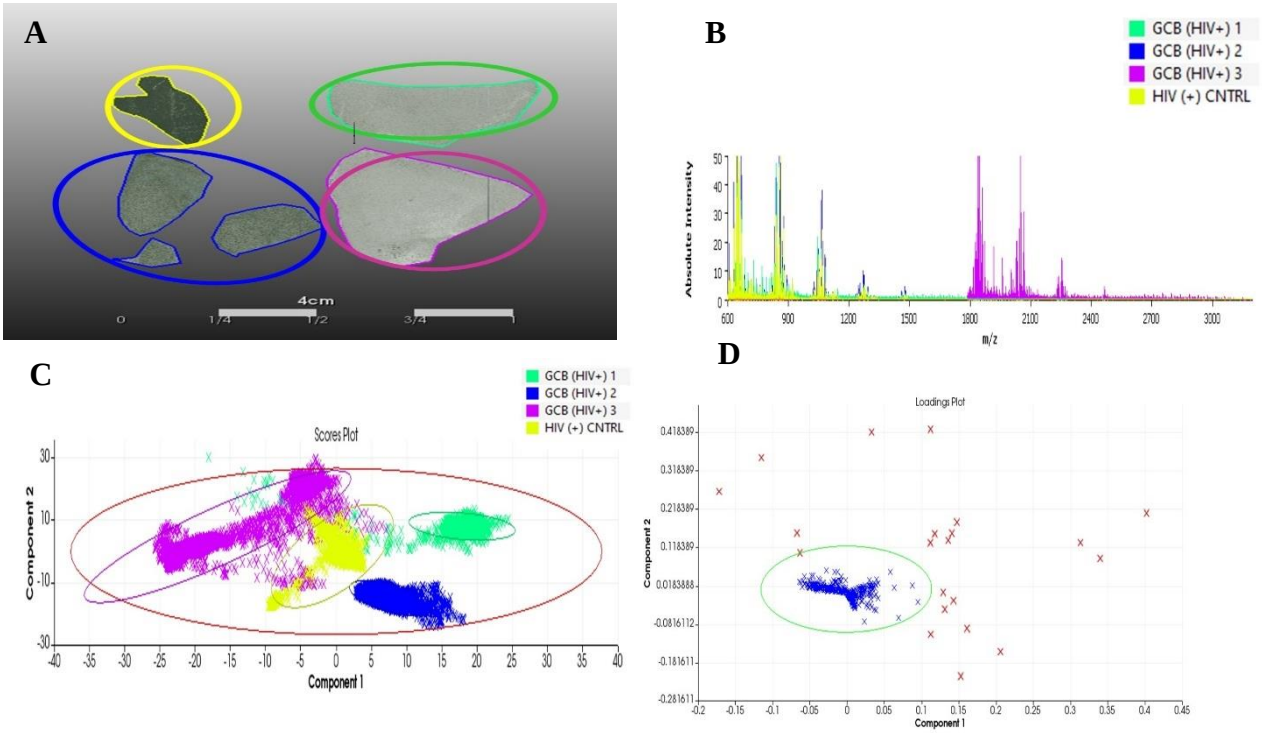


Figure 7. PCA analysis on GCB(HIV+) and (HIV+) control MADLI-IMS data set. (A) Stack view of scanned GCB (HIV+) cases (green, blue and purple) and (HIV+) control (Yellow). (B) Average spectra of the GCB(HIV+) cases and (HIV+) control clusters in the m/z range 606-3500. (C) PCA score plot indicated distinct clusters. (D) PCA loading plot indicated the contribution of each ion signal to the variance covered by PCA1 and PCA2.

4.2.2 PCA clustering between GCB DLBCL HIV(-) and HIV(-) control

The marked ROIs from three GCB DLBCL HIV (-) samples and HIV (-) control samples, were loaded onto SCiLs lab software (figure 8A). The average spectra of the ions were observed, and the peak picking function identified 364 ions in the range of m/z 600-3200 (figure 8B). The PCA score plot indicated three distinct clusters between the ion patterns of the mass spectra of the three GCB DLBCL HIV (-) samples and one HIV (-) control (figure 8C). Samples GCB HIV (-) 1 indicated in yellow, and GCB HIV (-) 2 indicated in green clustered together. The loading plot observed 9 different ion signals that contributed to the variance covered by PCA1 and PCA2 (figure 8D).

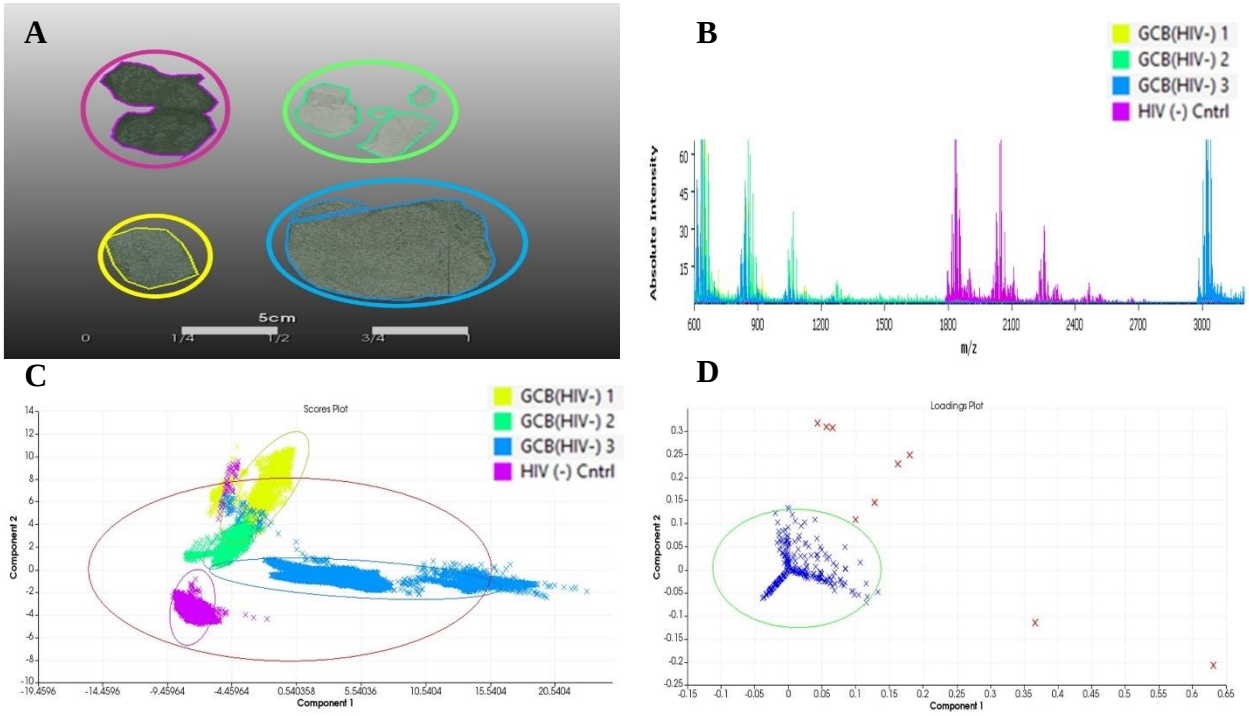


Figure 8. **PCA analysis on GCB(HIV-) and (HIV-) control MADLI-IMS data set.** (A) Stack view of Scanned GCB (HIV-) cases (yellow, green and blue) and (HIV-) control (purple). (B) Average spectra of the GCB(HIV-) cases and (HIV-) control clusters in the m/z range 600–3500 (C) Score (spectra) plotted against the first two components. (D) Loadings (m/z values) plotted against the first two components.

4.2.3 PCA analysis between ABC DLBCL HIV (+) and HIV (+) control

The marked ROIs from ABC DLBCL HIV (+) and HIV (+) control were loaded onto SCiLs lab software (figure 9A). The spectral peaks identified 335 ions in the range of m/z 607-1405 (figure 9B). The PCA score plot result indicated three clusters between the ion patterns of the mass spectra of the three ABC HIV (-) samples and HIV(-) control. The HIV (+) control (indicated in purple) and ABC DLBCL (HIV+) -2 (indicated in blue) clustered together (figure 9C). There were 11 different ion signals that contributed to the variance covered by PCA1 and PCA2, indicated on the loading plot (figure 9D).

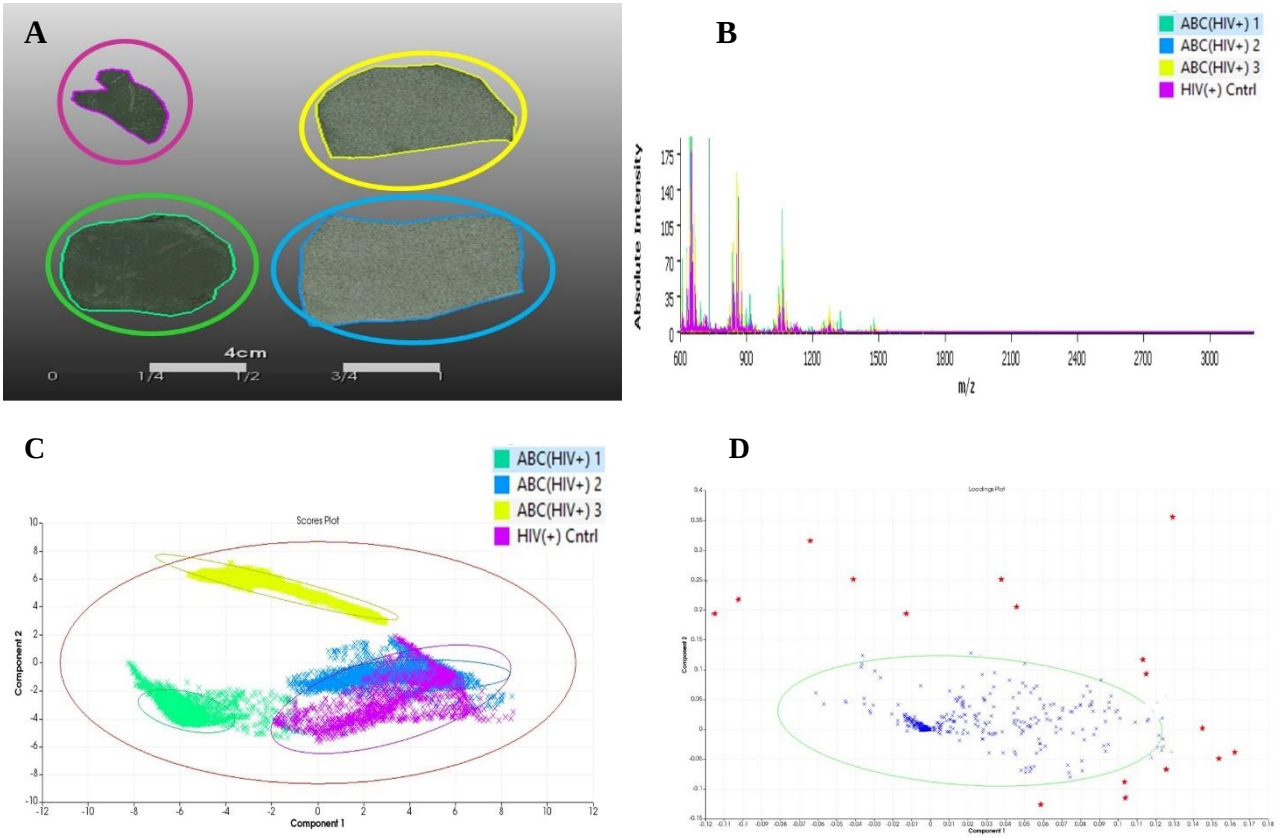


Figure 9. P**CA analysis on ABC(HIV+) and (HIV+) control MADLI-IMS data.** (A) Stack view of Scanned ABC (HIV+) cases (green, blue and yellow) and (HIV+) control (purple). (B) Average spectral peak of the ABC(HIV+) cases and (HIV+) control clusters in the m/z range 607–1405 (C) PCA score (spectra) plotted against components 1 and 2. (D) Loadings (m/z values) plotted against the components 1 and 2.

4.2.4 PCA analysis between ABC (HIV-) and HIV (-) control

The ROIs from three ABC DLBCL HIV(-) samples and one HIV (-) control sample were loaded onto SCiLs lab software (figure 10A). The spectral peaks identified 312 ions in the range of m/z 606-2519 (figure 10B). The PCA score plot showed similarities between ABC (HIV -) sample 1 (indicated in yellow) and ABC (HIV-) sample 2 (indicated in green) (figure 10C). A distinct separation was observed between the ion patterns of the mass spectra of the ABC (HIV-) samples 1, 2 and 3 (indicated in yellow, green, and purple respectively) and HIV (-) control (indicated in blue). The loading plot showed 9 different ion signals that contributed to the variance covered by PCA1 and PCA2 (figure 10D).

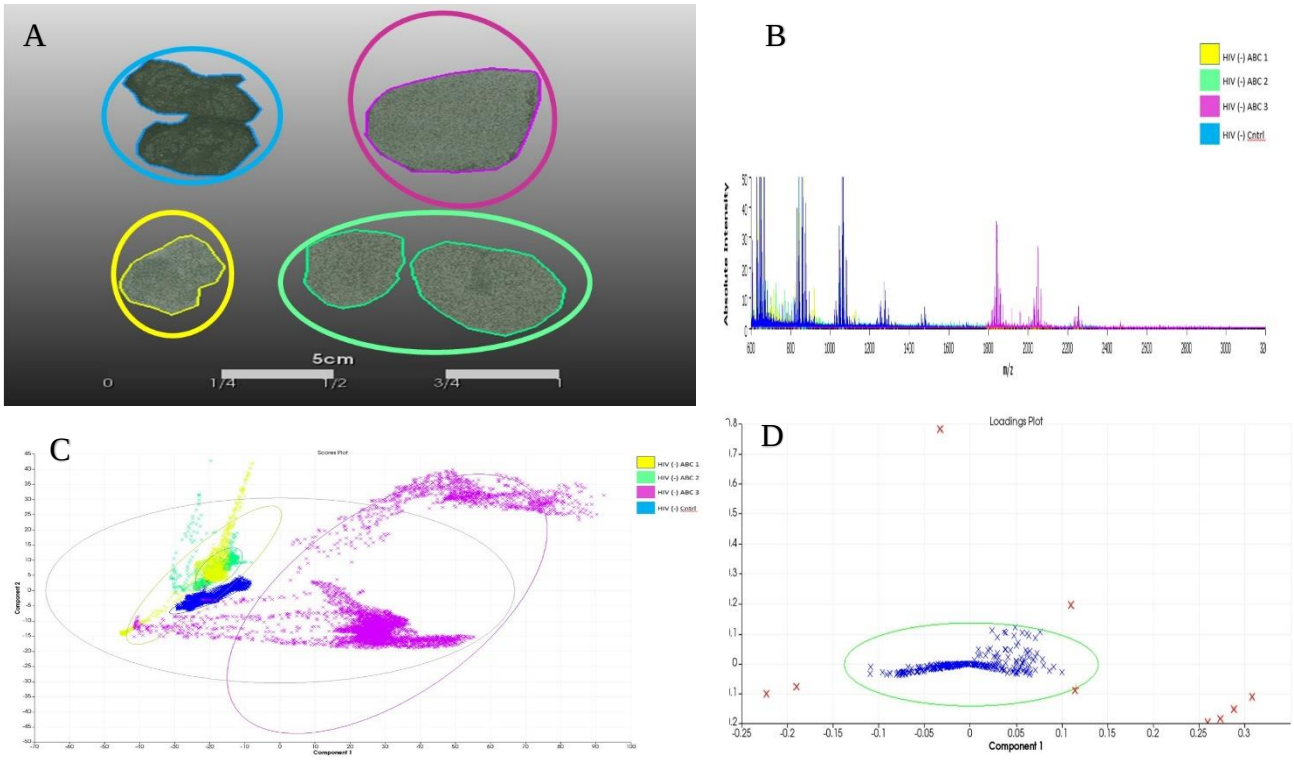


Figure 10. **PCA analysis on ABC(HIV-) and (HIV-) control MADLI-IMS data set**. (A) Stack view of scanned ABC (HIV-) samples (ABC sample1 yellow, ABC sample 2 green and ABC sample 3 purple) and (HIV-) control (blue). (B) The average spectral peak of the ABC(HIV-) cases and (HIV-) control m/x range 606-2519. (C) PCA score (spectra) plotted against component 1 and 2. (D) Loadings (m/z values) plotted against the component 1 and 2.

4.2.5 Exclusive ion mass (m/z) values identified in DLBCL cases

The peak detection function on the MALDI-IMS instrument identified 50 exclusive ion signals that distinguished DLBCL subtypes (appendix D, table 7). An online Venn diagram plotting software (http:/bioinformatics.psb.ugent.be), was used to visualize the exclusive ions among the different DLBCL subtypes. Exclusive ions found in each subtype were 11 ions for ABC HIV (+), in the range m/z 627-1030, 9 ions for ABC HIV (-), in the range m/z 643-2042, 21 ions for GCB HIV (+), in the range m/z 649-2086 and 9 ions for GCB HIV (-) in the range m/z 643-3025 (Figure 11).

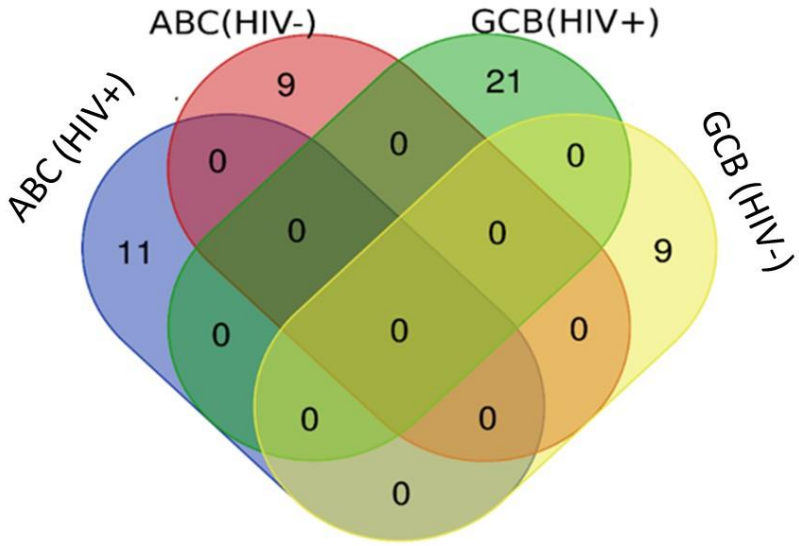


Figure 11. **Venn diagram showing a possible relationship of discovered ion signals among DLBCL cases**;[HIV (-) ABC and GCB; HIV (+) ABC and GCB]. As a result, none of the ion signals were common among the DLBCL cases.

We assessed the distribution of the ion with the highest intensity in the subtypes and the controls (Appendix C.2, figure 19). m/z 936.255 had the highest intensity in the HIV (-) control compared to the HIV (+) control and the DLBCL subtypes (Appendix C.2, figure 19 A,B,C). the AUC values was 0.936, this shows that the test analysis was able to identify m/z 936.255 was a true positive ion (Appendix C.2, figure 19 D).

4.3 LC-MS/MS data analysis

4.3.1 LC-MS/MS data distribution

FFPE tissues from 12 archived samples and 2 controls were subjected to LC-MS/MS based proteomic analysis. The Perseus software (version 1.6.10.43) was used to process and statistically analyse the data. After data processing, the protein entries reduced from 204 to 88. This meant that there were 88 proteins among all 14 DLBCL FFPE tissue specimens. All the potential contaminant proteins and reverse proteins were filtered out during data processing. An intensity box plot (figure 12) was generated to verify the distribution of the data. The protein intensities of each DLBCL subtype were not normally distributed, this was due to insufficient data and outliers of some protein intensities being extremely high or low in some samples. E.g. when actin cytoplasmic 1 was randomly selected (indicated in green), we observed an extremely high outlier in almost all the DLBCL samples except in ABCP2 and GCBP2 samples. The isoform2 of transcription intermediary factor 1 protein (indicated in red) was an extremely low outlier in samples ABCN1, ABCN3, ABCP1, and ABCP2. However, this protein was not present in other samples. Finally, the intensity of triosephosphate isomerase protein (indicated in blue) was an extremely low outlier in DLBCL samples except in GCBN2.

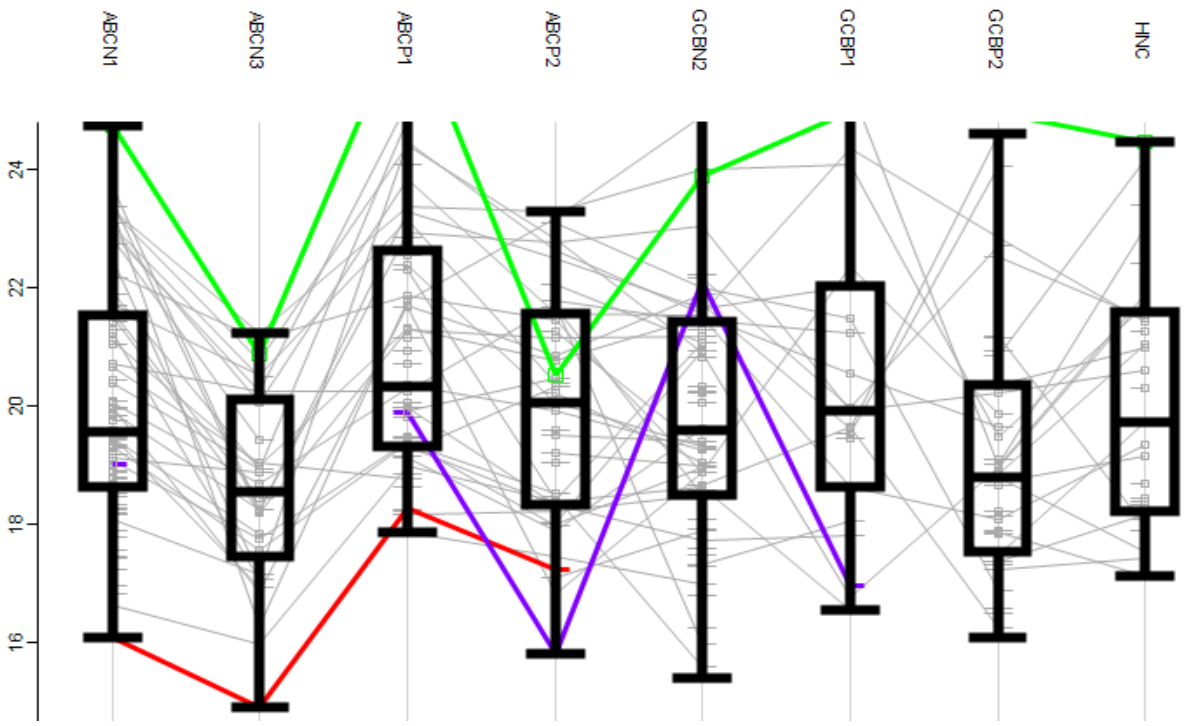


Figure 12. **Intensity Boxplot showing the spread of data after processing and normalization**. HIV negative DLBCL ABC subtype (ABCN1=sample 1 and ABCN3= sample 3). HIV positive DLBCL ABC subtype (ABCP1= sample 1 and ABCP2= sample 2). HIV negative DLBCL GCB subtype (GCBN2= sample 2). HIV positive GCB subtype (GCBP 1= sample 1 and GCB2= sample 2). HIV negative control (HNC) sample. Proteins intensities for Actin cytoplasmic 1 (green), Isoform2 of transcription intermediary factor 1 (red)and Isoform2 of triosephosphate isomerase (blue)

4.3.2 Experimental quality check

A multiple scatter plot (figure 13) was generated to determine if there was a linear correlation between the DLBCL samples and controls. The linear correlation was considered to be moderate if the Pearson correlation was (0.5< r <0.7). A strong linear correlation was considered to be strong if (r> 0.7) [69] and weak if (r<0.3). The overall linear correlation between variables was generally weak (r<0.3). There was a moderate positive linear relationship among samples (ABCN1 and ABCN3 ;r=0.588), (ABCP2 and ABCN1 r=0.562), (GCBP2 and ABCN1; r=0.653), (GCBP2 and ABCN3; r=0.525) and (GCBP2 and ABCP1; r=0.552).

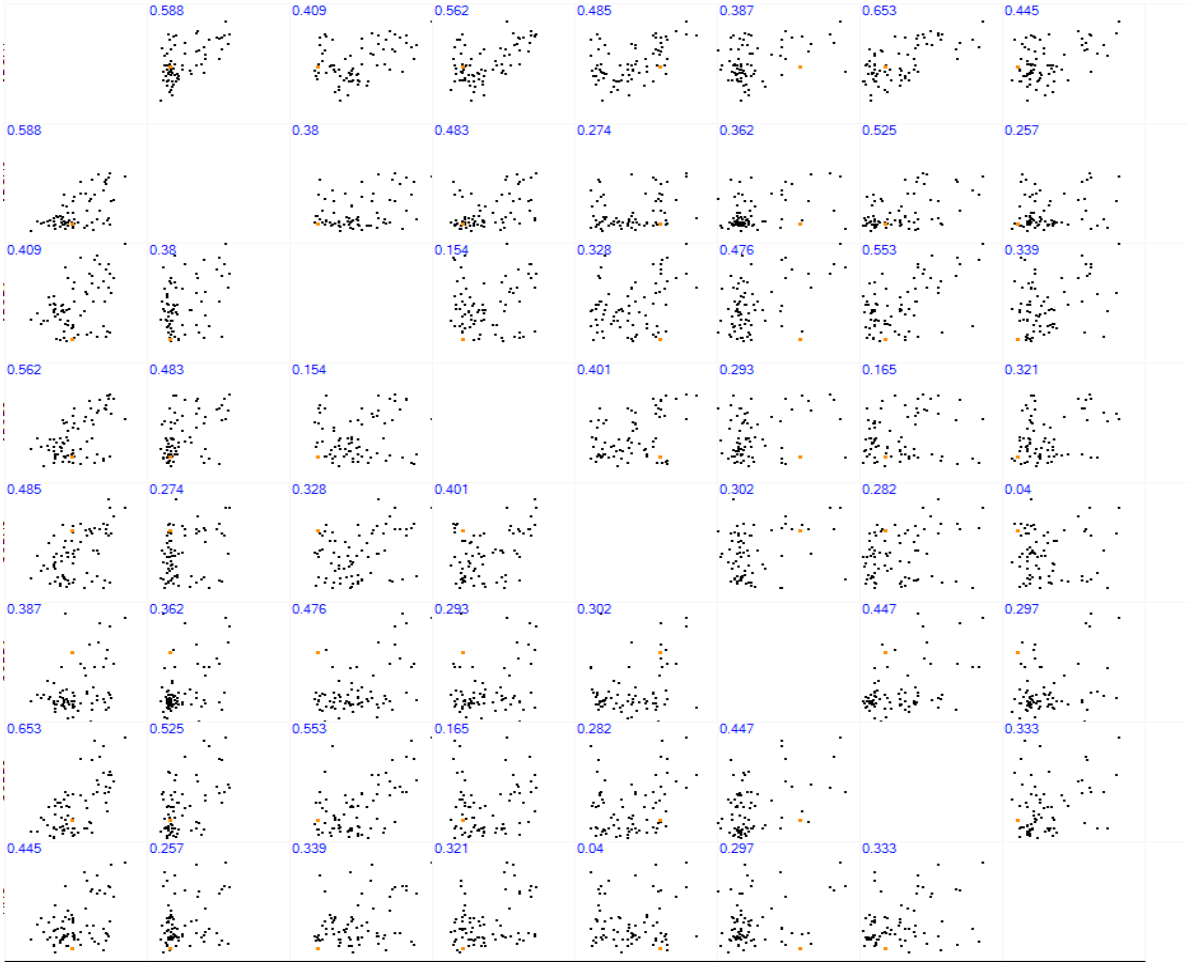


Figure 13: **Multiple Scatter plots illustrate the relationship between DLBCL cases and controls.** The correlation coefficient of (r < 0.3) and the lines are difficult to detect. While there was a moderate correlation between (ABCN1 and ABCN3 specimen ,r=0.588), (ABCP2 and ABCN1 specimen, r=0.562), (GCBP2 and ABCN1 specimen, r=0.653), (GCBP2 and ABCN3 specimen, r=0.525), and (GCBP2 and ABCP1, r=0.552).

4.3.3 Identification of potential biomarkers using Perseus

Taking each sample individually, we identified 88 and 32 proteins in samples ABCN1 and ABCN3, respectively. Similarly, the number of proteins in samples ABCP1 and ABCP2 were 52 and 45 respectively. Only 59 proteins were identified in the GCBN2 sample. A total of 21 and 41 proteins were identified in GCBP1 and GCBP2 samples, respectively (figure 14). In the HIV negative control (HNC), 31 proteins were identified, whilst no proteins were identified in samples ABCN2, ABCP3, GCBN1, GCBN3, and HPC. Further statistical analysis was carried out on identified proteins. Appendix J, table 10 summarizes the total number of proteins identified in this study.

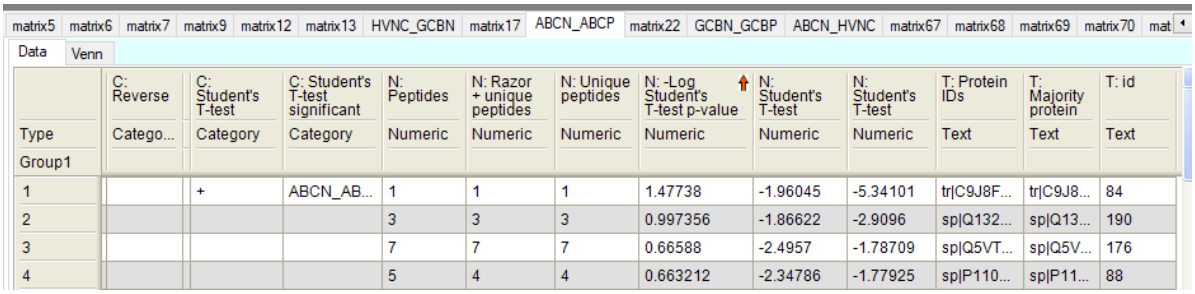
Total | Exclusive | Occurence

	Name	Total
1	ABCN1	88
2	ABCN3	32
3	ABCP1	52
4	ABCP2	45
5	GCBN2	59
6	GCBP1	21
7	GCBP2	41
8	HNC	31

Figure 1414. **The total number of proteins found in all the DLBCL subtypes.** HIV negative DLBCL ABC subtype (ABCN1=sample 1 and ABCN3= sample 3). HIV positive DLBCL ABC subtype (ABCP1= sample 1 and ABCP2= sample 2). HIV negative DLBCL GCB subtype (GCBN2= sample 2)

4.3.4 The identification of significantly differentially expressed proteins between HIV negative DLBCL ABC subtype (ABCN) and HIV positive DLBCL ABC (ABCP) subtypes.

A two-sample t-test (-log transformation P-values) at a threshold of 0.05 (5% probability error), identified human fructose-bisphosphate aldolase C (p value=1,47738) as the only differentially expressed proteins between ABCN and ABCP subtypes (figure 15). The KEGG pathway analysis revealed the involvement of fructose-bisphosphate aldolase C in the glycolytic pathway. A similar test was performed between the HVNC and ABCN specimens and there were no differentially expressed proteins.


matrix5 | matrix6 | matrix7 | matrix9 | matrix12 | matrix13 | HVNC_GCBN | matrix17 | ABCN_ABCP | matrix22 | GCBN_GCBP | ABCN_HVNC | matrix67 | matrix68 | matrix69 | matrix70 | mat

Data | Venn

	C: Reverse	C: Student's T-test	C: Student's T-test significant	N: Peptides	N: Razor + unique peptides	N: Unique peptides	N: -Log Student's T-test p-value	N: Student's T-test	N: Student's T-test	T: Protein IDs	T: Majority protein	T: id
Type	Catego...	Category	Category	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Text	Text	Text
Group1												
1		+	ABCN_AB...	1	1	1	1.47738	-1.96045	-5.34101	tr|C9J8F...	tr|C9J8...	84
2				3	3	3	0.997356	-1.86622	-2.9096	sp|Q132...	sp|Q13...	190
3				7	7	7	0.66588	-2.4957	-1.78709	sp|Q5VT...	sp|Q5V...	176
4				5	4	4	0.663212	-2.34786	-1.77925	sp|P110...	sp|P11...	88

Figure 15**. Two-sample t-test Statistical analysis** for ABCN and ABCP. The symbol '+' indicates statistical significance with respect to the specified threshold of 0.05 (5% probability error).

4.3.5 The identification of significantly differentially expressed proteins between HIV negative DLBCL GCB subtype (GCBN) and HIV positive DLBCL GCB (GCBP) subtypes.

A two-sample student t-test set at a threshold of 0.05 (5% probability error) was also applied to identify differentially expressed proteins between the GCBN and GCBP subtypes. There were no differentially expressed proteins between GCBN and GCBP subtypes (Appendix G, figure 22). A similar test was performed between the HVNC and GCBN specimens and there were no differentially expressed proteins.

4.3.6 Hierarchical clustering heatmap

The hierarchical clustering of proteins was done using the Euclidean distance method to visualize the differentially expressed protein in DLBCL subtypes and controls. The high and low protein expressions were indicated in red and green respectively (figure 16). We used a filtering strategy that required at least one protein to appear in 9 of the 14 specimens to be included for subsequent analyses. By comparing these proteins, we observed 6 differentially expressed among the DLBCL subtypes and controls (appendix E, table 8). These proteins were Tubulin alpha, ATP synthase, Enolase, Actin, Pyruvate kinase, and Ig kappa chain. Most of these proteins were part of the glycolytic pathway and ATP synthesis.

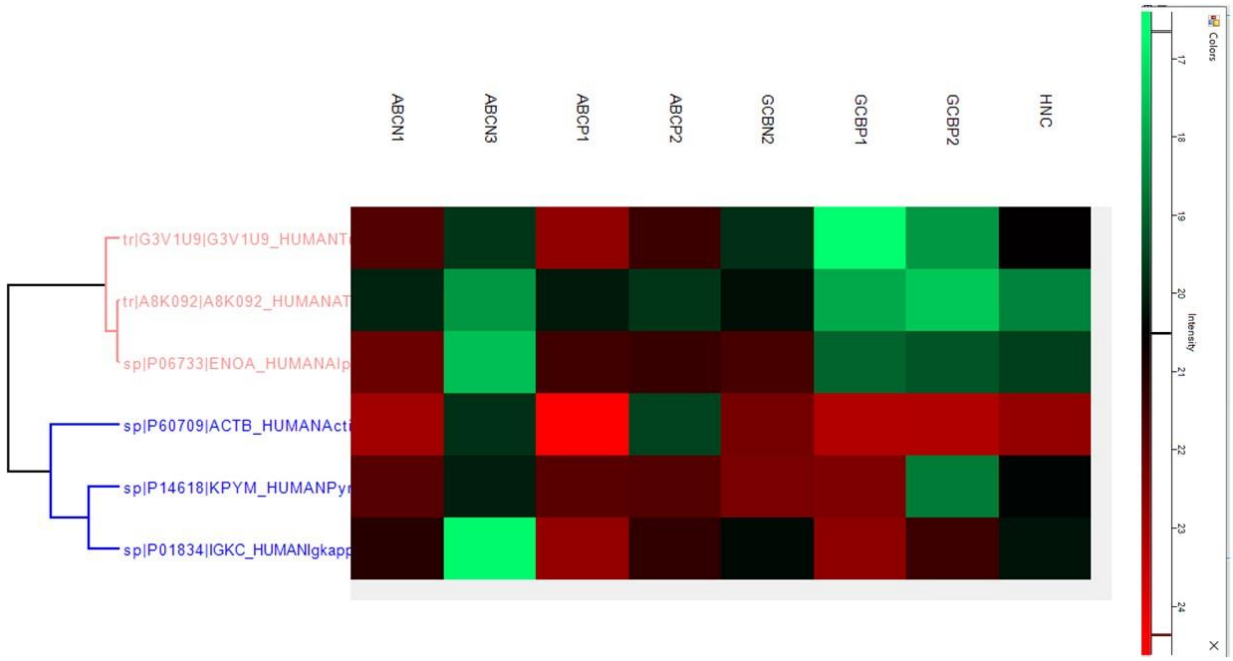


Figure 16. **Hierarchical clustering heatmap** of differentially expressed proteins among DLBCL cases and controls. The red indicates high expression and green indicate low expression.

Table 2. The regulation pattern of differentially expressed proteins

Protein name	**Regulation pattern**				
	ABC HIV (-)	ABC HIV(+)	GCB HIV (-)	GCB HIV (+)	HNC
Tubulin alpha	down	down	down	down	down
Pyruvate kinase	up	up	up	up	
ATP synthase	down	down	down	down	down
Actin	up	up	up	up	up
Ig Kappa Chain C	up	up	up	up	up
Alpha enolase	down	up	up	down	down

5. Discussion

FFPE tissue specimens are a valuable resource for translational proteomic studies because they provide archived patient samples, and well-documented clinical data (Table 1). Our findings regarding the patient's age showed a median age of 40 years for HIV positive DLBCL patients, which confirms previous studies that reported a median age of diagnosis for HIV-associated DLBCL to range from 30 to 49 years [122,123]. Although studies show an increase in the number of HIV positive females in South Africa, with studies showing predominance of HIV positive females, our study only had HIV negative females [124,125].

The PCA is the most commonly used statistical method used to visualize the overall variability of a data set (figure 7,8,9, and 10)[126]. The dimensionality of the date set is reduced to a 2D or 3D coordinate system, which uses dots to represent the mass spectrum. The spectra with similar features cluster together and the differences between sample groups are visualized in this coordinate system. The first two principal components (PCA1 and PCA2) in the 2D plot provides more than 80% of the total variance between the samples[126]. Using the 2D scores plot, we observed differences in both the HIV negative and positive GCB DLBCL samples (figure 7C and 8C). This suggested differences in the mass spectra ions between the diseased samples (DLBCL cases) and the normal tissue control (human tonsil). There was one sample in ABC DLBCL samples (both HIV positive and negative) that clustered together with the control sample (figure 9C and 10C). This clustering suggests that the ABC DLBCL samples had similar features to the control sample or that the largest source of variation among DLBCL cases and controls was similar. The loading plots of the PCA was used to visualize the ion signals with significant differences between the DLBCL samples and controls[126]. The significantly different ion signals were observed among the DLBCL samples and controls (figure 7D, 8D, 9D, and 10D). The significantly different ion signals were those that contributed to the variance covered by principle component 1 and component 2 in the PCA loading plots. To further validate this, we used the Venn diagram to visualize the distribution of the ion signal among the DLBCL subtypes, and there were 50 exclusive ion signals (figure 11). The number of exclusive ion signals observed in each of the DLBCL subtypes using a Venn diagram was similar to the total number of ion signals observed in the PCA loading plots.

FFPE tissue samples are not only used for the discovery of biomarkers but can also be used to understand molecular processes and pathways of cancer or other diseases[107]. We successfully applied LC-MS/MS to a total of 14 FFPE tissue samples and we identified a total of 88 proteins among all 14 FFPE tissue samples (Appendix E, table 8). The FFPE tissue processing method and LC-MS/MS strategies used herein have been performed in other published studies [84,122,123]. We used a filtering strategy that required at least one protein to appear in 9 of 14 specimens to be included for subsequent analyses. By comparing these proteins, there were 6 statistically significant proteins that were differentially expressed among the DLBCL subtypes and controls. These proteins were Tubulin alpha, ATP synthase, Enolase, Actin, Pyruvate kinase, and Ig kappa chain (figure 16). The majority of the proteins identified belonged to the glycolysis, ATP synthesis, and cellular movement. Glycolysis is a metabolic pathway by which cells breakdown glucose to generate their own energy in a form of ATP[127]. The production of ATP is essential for the synthesis of proteins. The synthesised proteins play a role in the regulation of cellular processes, and in providing cellular structure and movement [127,128].

Glycolytic proteins are usually overexpressed in cancer cells due to the increased cellular energy requirements during cell proliferation [129]. The glycolytic enzymes identified in our study were pyruvate kinase and alpha enolase. Pyruvate kinase is an enzyme involved in the last stages of glycolysis. It catalysis the formation of ATP and pyruvate from adenosine diphosphate (ADP), as phosphoenolpyruvate undergoes dephosphorylation[130]. The hallmark of cancer is increased glycolytic levels[81]. Previous studies observed an upregulation of pyruvate kinase in DLBCLs and other cancers [131–133], which is due to increased glycolysis. Warburg effect is the hallmark of cancer, whereby cancer cells undergo increased glycolysis to produce energy followed by lactic acid fermentation in the cytosol[132]. During these processes, pyruvate kinase is overexpressed and it catalyses the production of ATP during aerobic glycolysis [132]. Pyruvate kinase gives the tumour cells growth advantage and enables them to adapt and survive the tumour microenvironment[134].

Enolase is an enzymatic protein that catalyses the conversion of 2-phosphoglycerate (2-PG) to phosphoenolpyruvate (PEP)[130]. Various studies found enolase to play a role in metastasis and immune invasion by enhancing proteolytic activity through function as a plasminogen receptor[130,131]. In lung cancer, the upregulation of enolase was found to promote cell glycolysis, growth, migration, and invasion through FAK-mediated PI3K/AKT pathway[135]. Increased enolase expression was found to stimulate glycolysis, and this enhances resistance to

chemotherapy in gastric cancer patients. Therefore, this suggests the use of enolase as a biomarker to predict drug resistance and the overall survival of gastric cancer patients [136].

ATP synthase is an enzyme that plays a role in ATP synthesis and catalyses the oxidative phosphorylation during the respiratory process. We found ATP synthase to be downregulated in our study (table 2) which confirms previous findings from other studies [137]. Various studies found ATP synthase to be downregulated in almost all human cancers when compared with its expression in normal tissues[130,137]. Human cancers also upregulate the ATPase inhibitory factor 1 (IF1), which is the physiological inhibitor of the H+ -ATP synthase. The overexpression of IF1 reprograms energy metabolism to enhanced glycolysis by limiting ATP production by the H+-ATP synthase. Furthermore, the IF1-mediated inhibition of the H+ -ATP synthase promotes mitochondrial ROS, which regulates signalling pathways that favour cellular proliferation, activation of antioxidant defences, cell death resistance, and modulation of the tissue immune response, favouring the acquisition of several cancer traits [137].

In addition, we also found actin to be upregulated in our study, similar to other studies, that found actin as a key protein that provides cellular motility during the metastasis of cancer [138,139].

Tubulin is a protein that polymerized into filaments that form microtubules[140]. Microtubules are involved in cellular movement, intracellular trafficking, and mitosis. In cancer, the tubulin proteins are targets of the tubulin-binding chemotherapeutics, which supress the mitotic spindle to cause mitotic arrest and cell death. The over expression of different types of tubulin has been reported in a range of cancers. This expression is often correlated with poor prognosis and chemotherapy resistance in different types of cancers[141]. Tubulins and microtubules have been found to regulate mitochondrial metabolism. A recent study demonstrated that tubulin is capable of inter- acting with, and blocking the VDAC, thereby regulating ATP and metabolite compartmentalization and contributing to the War- burg effect[142].

Immunoglobulin kappa chain (IGKC) is a protein subunit of an antibody that is produced by immune cells. It links together with the heavy chain to form immunoglobulin (also known as antibodies)[87]. The IGKC is an immunological biomarker of prognosis and response to therapy in human cancers[143]. Little is known about the expression of IGKC in cancer cells. Various studies suggest that the secretion of IGKC by cancer cells has an unidentified capacity to promote cancer cell growth and survival[144].

Fructose-bisphosphate aldolase C (also known as aldolase C) was the only statistically different protein expressed between the ABC (HIV-) and ABC HIV (+) subtypes (p=1,47738). This

protein functions as an enzyme in glycolysis that splits aldol, fructose 1,6-bisphosphate, into the triose phosphates dihydroxyacetone phosphate (DHAP) and glyceraldehyde 3-phosphate (G3P) during glycolysis and gluconeogenesis [130]. Aldolase is differentially expressed in human tissues and has observed in different types of human cancers[130,132]. A previous study found aldolase enzymes to be upregulated during metastasis of Colorectal Adenocarcinoma[135]. Another study found aldolase enzymes especially aldolase A to be overexpressed in colon cancer[145]. Therefore, the expression of aldolase proteins is associated with an increased level of glycolysis and poor overall survival of cancer patients[146]. Aldolase has been validated though "-omics" database as a prognostic marker for different types of human cancers[147]. The clinical significance of aldolase C has not yet been studied in the context of DLBCL.

There were no differentially expressed proteins identified between both HIV negative and positive GCB DLBCL subtype. This may be caused by protein degradation during experiments[148].

Although we were able to successfully identify proteins using FFPE tissue, formalin fixation preparation conditions present a barrier that limits protein identification[107]. Protein-protein crosslinking occurs during formalin fixation. Chemically, during fixation formalin causes the methylation of amino acid side chains which forms a methylene bridge [70]. This may hinder the extraction of soluble proteins from FFPE tissue specimen [109,110]. Prolonged formalin fixation, results in increased protein-protein crosslinks which affect genomic and proteomic analysis[109,110]. Proteomic studies are limited by biological sample complexity, this may affect the analytical depth of the proteomic study[89]. In order to reduce sample complexity in our study, we used the FASP (Filter Aided sample preparation) method for FFPE sample preparation. The FASP protocol has been successfully applied by other researchers to FFPE tissue specimens and membrane proteins[93,118].

6. Conclusion

Using proteomic techniques, we identified and visualized differentially expressed protein in DLBCL subtypes and controls. The majority of these proteins belonged to glycolysis, ATP synthesis, and cellular movement. Fructose-bisphosphate aldolase C was the only significantly differentially expressed proteins between HIV negative ABC DLBCL and HIV positive ABC DLBCL subtypes. Suggesting its involvement in glycolysis which is a hallmark of cancer. Although, the aims and objectives of this study were archived. However, more research needs to study the clinicopathological correlation of HIV negative and HIV positive DLBCL with molecular data.

7. References

1. Plummer, M. *et al.* Global burden of cancers attributable to infections in 2012: a synthetic analysis. *Lancet Glob. Heal.* (2016). doi:10.1016/S2214-109X(16)30143-7

2. Ascierto, P. A. & Marincola, F. M. What have we learned from cancer immunotherapy in the last 3 years? *J. Transl. Med.* **12**, 1–11 (2014).

3. Aunan, J. R., Cho, W. C. & Søreide, K. The Biology of Aging and Cancer: A Brief Overview of Shared and Divergent Molecular Hallmarks. *Aging Dis.* **8**, 628–642 (2017).

4. Development and Spread of Cancer - Cancer - Merck Manuals Consumer Version. Available at: https://www.merckmanuals.com/home/cancer/overview-of-cancer/development-and-spread-of-cancer. (Accessed: 20th February 2018)

5. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).

6. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).

7. Tainsky, M. A. Genomic and proteomic biomarkers for cancer: a multitude of opportunities. *Biochim. Biophys. Acta* **1796**, 176–93 (2009).

8. WHO | World Cancer Report 2014. *WHO* (2015).

9. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2017. *CA. Cancer J. Clin.* **67**, 7–30 (2017).

10. De Felice, F. *et al.* Head and neck diffuse large B cell lymphomas (HN-DLBCL) in human immunodeficiency virus (HIV) positive patients: long-term results in the highly active antiretroviral therapy (HAART) era. *Eur. Arch. Oto-Rhino-Laryngology* **274**, 3735–3739 (2017).

11. *DIFFUSE LARGE B-CELL LYMPHOMA Union for International Cancer Control 2014 Review of Cancer Medicines on the WHO List of Essential Medicines DIFFUSE LARGE B-CELL LYMPHOMA Executive Summary*.

12. De Felice, F. *et al.* Head and neck diffuse large B cell lymphomas (HN-DLBCL) in

human immunodeficiency virus (HIV) positive patients: long-term results in the highly active antiretroviral therapy (HAART) era. *Eur. Arch. Oto-Rhino-Laryngology* **274**, 3735–3739 (2017).

13. Sehn, L. H. & Gascoyne, R. D. Diffuse large B-cell lymphoma: optimizing outcome in the context of clinical and biologic heterogeneity. *Blood* **125**, 22–32 (2015).

14. Pasqualucci, L. & Dalla-Favera, R. The Genetic Landscape of Diffuse Large B-Cell Lymphoma. *Semin. Hematol.* **52**, 67–76 (2015).

15. Coiffier, B. Diffuse large cell lymphoma. *Curr. Opin. Oncol.* **13**, 325–34 (2001).

16. Dunleavy, K. & Wilson, W. H. How I treat HIV-associated lymphoma. *Blood* **119**, 3245–3255 (2012).

17. Campo, E. *et al.* The 2008 WHO classification of lymphoid neoplasms and beyond: Evolving concepts and practical applications. *Blood* **117**, 5019–5032 (2011).

18. Chao, C. *et al.* Epstein-Barr Virus Infection and Expression of B-cell Oncogenic Markers in HIV-Related Diffuse Large B-cell Lymphoma. *Clin. Cancer Res.* **18**, 4702–4712 (2012).

19. Song, C.-G. *et al.* Epstein-Barr Virus-Positive Diffuse Large B-Cell Lymphoma in the Elderly: A Matched Case-Control Analysis. *PLoS One* **10**, e0133973 (2015).

20. Herndier, B. G. *et al.* Acquired immunodeficiency syndrome-associated T-cell lymphoma: evidence for human immunodeficiency virus type 1-associated T-cell transformation. *Blood* **79**, 1768–74 (1992).

21. Grogg, K. L., Miller, R. F. & Dogan, A. HIV infection and lymphoma. *Journal of Clinical Pathology* **60**, 1365–1372 (2007).

22. Bibas, M. & Antinori, A. EBV and HIV-Related Lymphoma. *Mediterr. J. Hematol. Infect. Dis.* (2009). doi:10.4084/mjhid.2009.032

23. Besson, C. *et al.* Changes in AIDS-related lymphoma since the era of highly active antiretroviral therapy. *Blood* **98**, 2339–44 (2001).

24. Besson, C. *et al.* Outcomes for HIV-associated diffuse large B-cell lymphoma in the modern combined antiretroviral therapy era. *Aids* **31**, 2493–2501 (2017).

25. Knowles, D. M. Etiology and pathogenesis of AIDS-related non-Hodgkin's lymphoma.

Hematol. Oncol. Clin. North Am. **17**, 785–820 (2003).

26. Blinder, V. S., Chadburn, A., Furman, R. R., Mathew, S. & Leonard, J. P. Improving outcomes for patients with Burkitt lymphoma and HIV. *AIDS Patient Care STDS* **22**, 175–87 (2008).

27. Rinaldi, A., Capello, D., Zucca, E., Gaidano, G. & Bertoni, F. Genome-Wide DNA Profiling of HIV-Related B-Cell Lymphomas. in *Methods in molecular biology (Clifton, N.J.)* **973**, 213–226 (2013).

28. Beral, V., Peterman, T., Berkelman, R. & Jaffe, H. AIDS-associated non-Hodgkin lymphoma. *Lancet (London, England)* **337**, 805–9 (1991).

29. Landgren, O. *et al.* Circulating Serum Free Light Chains As Predictive Markers of AIDS-Related Lymphoma. *J. Clin. Oncol.* **28**, 773–779 (2010).

30. Swerdlow, S. H. *et al.* The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* **127**, 2375–2390 (2016).

31. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).

32. Dunleavy, K., Roschewski, M. & Wilson, W. H. Precision Treatment of Distinct Molecular Subtypes of Diffuse Large B-cell Lymphoma: Ascribing Treatment Based on the Molecular Phenotype. *Clin. Cancer Res.* **20**, 5182–5193 (2014).

33. Alizadeh, a a *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–11 (2000).

34. Nowakowski, G. S. & Czuczman, M. S. ABC, GCB, and Double-Hit Diffuse Large B-Cell Lymphoma: Does Subtype Make a Difference in Therapy Selection? *Am. Soc. Clin. Oncol. Educ. B.* **35**, e449–e457 (2015).

35. Dunleavy, K. & Wilson, W. H. Appropriate management of molecular subtypes of diffuse large B-cell lymphoma. *Oncology (Williston Park).* **28**, 326–34 (2014).

36. Basso, K. & Dalla-Favera, R. Germinal centres and B cell lymphomagenesis. *Nat. Rev. Immunol.* **15**, 172–184 (2015).

37. Pasqualucci, L. & Dalla-Favera, R. The Genetic Landscape of Diffuse Large B-Cell Lymphoma. *Semin. Hematol.* **52**, 67–76 (2015).

38. Li, S., Young, K. H. & Medeiros, L. J. Diffuse large B-cell lymphoma. *Pathology* **50**, 74–87 (2018).

39. Sehn, L. H. *et al.* Diffuse large B-cell lymphoma: optimizing outcome in the context of clinical and biologic heterogeneity. *Blood* **125**, 22–32 (2015).

40. Li, S. *et al.* BCL6 rearrangement indicates poor prognosis in diffuse large B-cell lymphoma patients: A meta-analysis of cohort studies. *J. Cancer* **10**, 530–538 (2019).

41. Carbone, A., Gloghini, A., Kwong, Y.-L. & Younes, A. Diffuse large B cell lymphoma: using pathologic and molecular biomarkers to define subgroups for novel therapy. *Ann. Hematol.* **93**, 1263–77 (2014).

42. Akyurek, N., Uner, A., Benekli, M. & Barista, I. Prognostic significance of *MYC* , *BCL2* , and *BCL6* rearrangements in patients with diffuse large B-cell lymphoma treated with cyclophosphamide, doxorubicin, vincristine, and prednisone plus rituximab. *Cancer* **118**, 4173–4183 (2012).

43. Blenk, S. *et al.* Germinal center B cell-like (GCB) and activated B cell-like (ABC) type of diffuse large B cell lymphoma (DLBCL): analysis of molecular predictors, signatures, cell cycle state and patient survival. *Cancer Inform.* **3**, 399–420 (2007).

44. Turvey, S. E. *et al.* The CARD11-BCL10-MALT1 (CBM) signalosome complex: Stepping into the limelight of human primary immunodeficiency. *J. Allergy Clin. Immunol.* **134**, 276–284 (2014).

45. Carbone, A., Gloghini, A., Kwong, Y.-L. & Younes, A. Diffuse large B cell lymphoma: using pathologic and molecular biomarkers to define subgroups for novel therapy. *Ann. Hematol.* **93**, 1263–77 (2014).

46. Quintanilla-Martinez, L. The 2016 updated WHO classification of lymphoid neoplasias. *Hematol. Oncol.* **35**, 37–45 (2017).

47. Hu, S. *et al.* MYC/BCL2 protein coexpression contributes to the inferior survival of activated B-cell subtype of diffuse large B-cell lymphoma and demonstrates high-risk gene expression signatures: A report from the International DLBCL Rituximab-CHOP Consortium Program. *Blood* **121**, 4021–4031 (2013).

48. Nicolae, A. *et al.* EBV-positive large B-cell lymphomas in young patients: A nodal lymphoma with evidence for a tolerogenic immune environment. *Blood* **126**, 863–872

(2015).

49. Adam, P., Bonzheim, I., Fend, F. & Quintanilla-Martínez, L. Epstein-Barr Virus-positive Diffuse Large B-cell Lymphomas of the Elderly. *Adv. Anat. Pathol.* **18**, 349–355 (2011).

50. Gonzalez-Farre, B. *et al.* Burkitt-like lymphoma with 11q aberration: A germinal center-derived lymphoma genetically unrelated to Burkitt lymphoma. *Haematologica* **104**, 1822–1829 (2019).

51. Salaverria, I. *et al.* A recurrent 11q aberration pattern characterizes a subset of MYC-negative high-grade B-cell lymphomas resembling Burkitt lymphoma. *Blood* **123**, 1187–1198 (2014).

52. Liu, Y. & Barta, S. K. Diffuse large B-cell lymphoma: 2019 update on diagnosis, risk stratification, and treatment. *Am. J. Hematol.* **94**, 604–616 (2019).

53. Liu, Y. & Barta, S. K. Diffuse large B-cell lymphoma: 2019 update on diagnosis, risk stratification, and treatment. *American Journal of Hematology* **94**, 604–616 (2019).

54. Carbone, A. *et al.* Diagnosis and management of lymphomas and other cancers in HIV-infected patients. *Nature Reviews Clinical Oncology* **11**, 223–238 (2014).

55. Choi, W. W. L. *et al.* A new immunostain algorithm classifies diffuse large B-cell lymphoma into molecular subtypes with high accuracy. *Clin. Cancer Res.* **15**, 5494–5502 (2009).

56. Shustik, J. *et al.* Correlations between BCL6 rearrangement and outcome in patients with diffuse large B-cell lymphoma treated with CHOP or R-CHOP. *Haematologica* **95**, 96–101 (2010).

57. Cormier, J. N. & Pollock, R. E. Soft tissue sarcomas. *CA. Cancer J. Clin.* **54**, 94–109

58. Westin, J. R. & Fayad, L. E. Beyond R-CHOP and the IPI in large-cell lymphoma: Molecular markers as an opportunity for stratification. *Curr. Hematol. Malig. Rep.* **4**, 218–224 (2009).

59. Nowakowski, G. S. *et al.* Beyond RCHOP: A Blueprint for Diffuse Large B Cell Lymphoma Research. *J. Natl. Cancer Inst.* **108**, djw257 (2016).

60. Diffuse Large B-Cell Lymphoma: Treatment Options - LRF. Available at:

https://lymphoma.org/aboutlymphoma/nhl/dlbcl/dlbcltreatment/. (Accessed: 15th January 2020)

61. Chiappella, A., Santambrogio, E., Castellino, A., Nicolosi, M. & Vitolo, U. Integrating novel drugs to chemoimmunotherapy in diffuse large B-cell lymphoma. *Expert Rev. Hematol.* **10**, 697–705 (2017).

62. Pfreundschuh, M. [Current therapeutic strategies for diffuse large B-cell lymphoma]. *Internist (Berl).* **57**, 214–21 (2016).

63. Lossos, I. S. & Morgensztern, D. Prognostic Biomarkers in Diffuse Large B-Cell Lymphoma. *J. Clin. Oncol.* **24**, 995–1007 (2006).

64. Zhou, Z. *et al.* An enhanced International Prognostic Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era. *Blood* **123**, 837–842 (2014).

65. Biomarkers - an overview | ScienceDirect Topics. Available at: https://www.sciencedirect.com/topics/neuroscience/biomarkers. (Accessed: 23rd May 2018)

66. Tainsky, M. A. Genomic and proteomic biomarkers for cancer: A multitude of opportunities. *Biochim. Biophys. Acta - Rev. Cancer* **1796**, 176–193 (2009).

67. Immunohistochemical overexpression of BCL-2 protein predicts an inferior survival in patients with primary central nervous system diffuse large B-c... - PubMed - NCBI. Available at: https://www.ncbi.nlm.nih.gov/pubmed/31702637/. (Accessed: 21st December 2019)

68. Perry, A. M. *et al.* MYC and BCL2 protein expression predicts survival in patients with diffuse large B-cell lymphoma treated with rituximab. *Br. J. Haematol.* **165**, 382–391 (2014).

69. Tyanova, S. & Cox, J. Chapter 7. **1711**, 133–148

70. Iqbal, J. *et al. BCL2* Expression Is a Prognostic Marker for the Activated B-Cell–Like Type of Diffuse Large B-Cell Lymphoma. *J. Clin. Oncol.* **24**, 961–968 (2006).

71. Tsuyama, N. *et al.* BCL2 expression in DLBCL: reappraisal of immunohistochemistry with new criteria for therapeutic biomarker evaluation. *Blood* **130**, 489–500 (2017).

72. Wilson, W. H. *et al.* Relationship of p53, bcl-2, and tumor proliferation to clinical drug resistance in non-Hodgkin's lymphomas. *Blood* **89**, 601–9 (1997).

73. Tang, Y. L., Zhou, Y., Cheng, L. L., Su, Y. Z. & Wang, C. Bin. BCL2/Ki-67 index predict survival in germinal center B-cell-like diffuse large B-cell lymphoma. *Oncol. Lett.* **14**, 3767–3773 (2017).

74. He, X. *et al.* Ki-67 is a valuable prognostic predictor of lymphoma but its utility varies in lymphoma subtypes: evidence from a systematic meta-analysis. *BMC Cancer* **14**, 153 (2014).

75. Brown, D. C. & Gatter, K. C. Ki67 protein: the immaculate deception? *Histopathology* **40**, 2–11 (2002).

76. BCL2/Ki-67 index predict survival in germinal center B-cell-like diffuse large B-cell lymphoma. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5588076/. (Accessed: 20th December 2019)

77. Endl, E. & Gerdes, J. The Ki-67 Protein: Fascinating Forms and an Unknown Function. *Exp. Cell Res.* **257**, 231–237 (2000).

78. Broyde, A. *et al.* Role and prognostic significance of the Ki-67 index in non-Hodgkin's lymphoma. *Am. J. Hematol.* **84**, 338–343 (2009).

79. Hinkle, C., Makar, G. S., Brody, J. D., Ahmad, N. & Zhu, G. G. HIV-Associated "Double-Hit" Lymphoma of the Tonsil: A First Reported Case. *Head Neck Pathol.* (2020). doi:10.1007/s12105-020-01135-1

80. Mutational profile and prognostic significance of TP53 in diffuse large B-cell lymphoma patients treated with R-CHOP: report from an International ... - PubMed - NCBI. Available at: https://www.ncbi.nlm.nih.gov/pubmed/22955915/. (Accessed: 23rd December 2019)

81. Fouad, Y. A. & Aanei, C. Revisiting the hallmarks of cancer. *Am. J. Cancer Res.* **7**, 1016–1036 (2017).

82. Voropaeva, E. N., Voevoda, M. I., Pospelova, T. I. & Maksimov, V. N. Prognostic impact of the *TP53* rs1625895 polymorphism in DLBCL patients. *Br. J. Haematol.* **169**, 32–35 (2015).

83. Zainuddin, N. *et al.* TP53 mutations predict for poor survival in de novo diffuse large B-cell lymphoma of germinal center subtype. *Leuk. Res.* **33**, 60–66 (2009).

84. Xu-Monette, Z. Y. *et al.* Mutational profile and prognostic significance of TP53 in diffuse large B-cell lymphoma patients treated with R-CHOP: Report from an International DLBCL Rituximab-CHOP Consortium Program Study. *Blood* **120**, 3986–3996 (2012).

85. Harris, C. C. Structure and function of the p53 tumor suppressor gene: clues for rational cancer therapeutic strategies. *J. Natl. Cancer Inst.* **88**, 1442–55 (1996).

86. Jerkeman, M. *et al.* Prognostic implications of BCL6 rearrangement in uniformly treated patients with diffuse large B-cell lymphoma--a Nordic Lymphoma Group study. *Int. J. Oncol.* **20**, 161–5 (2002).

87. Tomita, N. *et al.* Diffuse large B cell lymphoma without immunoglobulin light chain restriction by flow cytometry. *Acta Haematol.* **121**, 196–201 (2009).

88. Tuma, J. M., & Pratt, J. M. (1982). Clinical child psychology practice and training: A survey. \ldots of Clinical Child & Adolescent Psychology, 137(August 2012), 37–41. http://doi.org/10.1037/a0022390 *et al.* Initial sequencing and analysis of the human genome. *Nature* (2001). doi:10.1038/35057062

89. Chandramouli, K. & Qian, P.-Y. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Hum. Genomics Proteomics* **2009**, (2009).

90. Hixson, K. K., Lopez-Ferrer, D., Robinson, E. W. & Paša-Tolić, L. Proteomics. *Encycl. Spectrosc. Spectrom.* 766–773 (2017). doi:10.1016/B978-0-12-803224-4.00061-3

91. Paulo, J. A., Lee, L. S., Banks, P. A., Steen, H. & Conwell, D. L. Proteomic analysis of formalin-fixed paraffin-embedded pancreatic tissue using liquid chromatography tandem mass spectrometry. *Pancreas* **41**, 175–85 (2012).

92. Yu, L.-R., Stewart, N. A. & Veenstra, T. D. Proteomics: The Deciphering of the Functional Genome. *Essentials Genomic Pers. Med.* 89–96 (2010). doi:10.1016/B978-0-12-374934-5.00008-8

93. Wiśniewski, J. R. Proteomic sample preparation from formalin fixed and paraffin embedded tissue. *J. Vis. Exp.* (2013). doi:10.3791/50589

94. Paulo, J. A., Kadiyala, V., Banks, P. A., Steen, H. & Conwell, D. L. Mass spectrometry-based proteomics for translational research: a technical overview. *Yale J. Biol. Med.* **85**, 59–73 (2012).

95. Zhang, H. & Ge, Y. Comprehensive analysis of protein modifications by top-down mass spectrometry. *Circ. Cardiovasc. Genet.* **4**, 711 (2011).

96. Gundry, R. L. *et al.* Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Curr. Protoc. Mol. Biol.* **Chapter 10**, Unit10.25 (2009).

97. Alexandrov, T. MALDI imaging mass spectrometry: statistical data analysis and current computational challenges. *BMC Bioinformatics* (2012). doi:10.1186/1471-2105-13-s16-s11

98. Franck, J. *et al.* MALDI imaging mass spectrometry: state of the art technology in clinical proteomics. *Mol. Cell. Proteomics* **8**, 2023–33 (2009).

99. Aburaya, S., Aoki, W., Minakuchi, H. & Ueda, M. Definitive screening design enables optimization of LC–ESI–MS/MS parameters in proteomics. *Biosci. Biotechnol. Biochem.* **81**, 2237–2243 (2017).

100. Balluff, B., Schöne, C., Höfler, H. & Walch, A. MALDI imaging mass spectrometry for direct tissue analysis: Technological advancements and recent applications. *Histochemistry and Cell Biology* **136**, 227–244 (2011).

101. Alexandrov, T. *MALDI imaging mass spectrometry: statistical data analysis and current computational challenges*. (2012). doi:10.1186/1471-2105-13-S16-S11

102. Plechawska-Wojcik, M. *4 A Comprehensive Analysis of MALDI-TOF Spectrometry Data*.

103. Gessel, M. M., Norris, J. L. & Caprioli, R. M. MALDI imaging mass spectrometry: Spatial molecular analysis to enable a new age of discovery. *J. Proteomics* **107**, 71–82 (2014).

104. Gorzolka, K. & Walch, A. MALDI mass spectrometry imaging of formalin-fixed paraffin-embedded tissues in clinical research. *Histol. Histopathol.* **29**, 1365–76 (2014).

105. Caprioli, R. M., Farmer, T. B. & Gile, J. Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. *Anal. Chem.* **69**, 4751–4760 (1997).

106. Chaurand, P., Schwartz, S. A. & Caprioli, R. M. Peer Reviewed: Profiling and Imaging Proteins in Tissue Sections by MS. *Anal. Chem.* **76**, 86 A-93 A (2004).

107. Magdeldin, S. & Yamamoto, T. Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues. *Proteomics* **12**, 1045–58 (2012).

108. Föll, M. C. *et al.* Reproducible proteomics sample preparation for single FFPE tissue slices using acid-labile surfactant and direct trypsinization. *Clin. Proteomics* **15**, 11 (2018).

109. Wiśniewski, J. R. Proteomic sample preparation from formalin fixed and paraffin embedded tissue. *J. Vis. Exp.* (2013). doi:10.3791/50589

110. Casadonte, R. & Caprioli, R. M. Proteomic analysis of formalin-fixed paraffin-embedded tissue. *Nat. Protoc.* **6**, 1695–709 (2011).

111. Fend, F. & Raffeld, M. Laser capture microdissection in pathology. *J. Clin. Pathol.* **53**, 666–72 (2000).

112. Proteomics | Nature Biotechnology. Available at: https://www.nature.com/articles/s41587-019-0276-y. (Accessed: 13th January 2020)

113. Walch, A., Rauser, S., Deininger, S.-O. & Höfler, H. MALDI imaging mass spectrometry for direct tissue analysis: a new frontier for molecular histology. *Histochem. Cell Biol.* **130**, 421–34 (2008).

114. Paulo, J. A. *et al.* A proteomic comparison of formalin-fixed paraffin-embedded pancreatic tissue from autoimmune pancreatitis, chronic pancreatitis, and pancreatic cancer. *JOP* **14**, 405–14 (2013).

115. *SCiLS Lab 2D: Comparative Analysis for Uncovering Discriminative M/z-markers*.

116. Klein, O. *et al.* MALDI imaging mass spectrometry: Discrimination of pathophysiological regions in traumatized skeletal muscle by characteristic peptide signatures. *Proteomics* **14**, 2249–2260 (2014).

117. Sun, C. S. & Markey, M. K. Recent advances in computational analysis of mass

spectrometry for proteomic profiling. *Journal of Mass Spectrometry* **46**, 443–456 (2011).

118. FFPE-FASP | Max Planck Institute of Biochemistry. Available at: https://www.biochem.mpg.de/1143035/FFPE-FASP. (Accessed: 4th February 2019)

119. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–70 (2003).

120. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

121. Tyanova, S. & Cox, J. Perseus: A bioinformatics platform for integrative analysis of proteomics data in cancer research. in *Methods in Molecular Biology* **1711**, 133–148 (Humana Press Inc., 2018).

122. Baptista, M. J. *et al.* HIV-infection impact on clinical-biological features and outcome of diffuse large B-cell lymphoma treated with R-CHOP in the combination antiretroviral therapy era. *AIDS* **29**, 811–818 (2015).

123. Wiggill, T. M., Mantina, H., Willem, P., Perner, Y. & Stevens, W. S. Changing pattern of lymphoma subgroups at a tertiary academic complex in a high-prevalence HIV setting: A South African perspective. *J. Acquir. Immune Defic. Syndr.* **56**, 460–466 (2011).

124. Shisana, O. *et al.* Does marital status matter in an HIV hyperendemic country? Findings from the 2012 South African National HIV Prevalence, Incidence and Behaviour Survey. *AIDS Care - Psychol. Socio-Medical Asp. AIDS/HIV* **28**, 234–241 (2016).

125. UNAIDS Prevention gap report 2016 | Children & AIDS. Available at: https://www.childrenandaids.org/node/793. (Accessed: 23rd July 2020)

126. Cho, Y. T., Chiang, Y. Y., Shiea, J. & Hou, M. F. Combining MALDI-TOF and molecular imaging with principal component analysis for biomarker discovery and clinical diagnosis of cancer. *Genomic Med. Biomarkers, Heal. Sci.* **4**, 3–6 (2012).

127. Glycolysis | definition of glycolysis by Medical dictionary. Available at:

https://medical-dictionary.thefreedictionary.com/glycolysis. (Accessed: 5th February 2020)

128. What Is Protein Synthesis - Protein Synthesis. Available at: https://www.proteinsynthesis.org/what-is-protein-synthesis/. (Accessed: 5th February 2020)

129. Yu, L., Chen, X., Sun, X., Wang, L. & Chen, S. J o u r n a l o f C a n c e r The Glycolytic Switch in Tumors : How Many Players Are Involved ? **8**, (2017).

130. Li, X., Gu, J. & Zhou, Q. Review of aerobic glycolysis and its key enzymes – new targets for lung cancer therapy. **6**, 17–24 (2015).

131. Ludvigsen, M. *et al.* Proteomic profiling identifies outcome-predictive markers in patients with peripheral T-cell lymphoma, not otherwise specified. *Blood Adv.* **2**, 2533–2542 (2018).

132. Epstein, T., Gatenby, R. A. & Brown, J. S. The Warburg effect as an adaptation of cancer cells to rapid fluctuations in energy demand. 1–14 (2017).

133. Magangane, P., Sookhayi, R., Govender, D. & Naidoo, R. Determining protein biomarkers for DLBCL using FFPE tissues from HIV negative and HIV positive patients. *J. Mol. Histol.* **47**, 565–577 (2016).

134. Guo, C. *et al.* Tumor pyruvate kinase M2: A promising molecular target of gastrointestinal cancer. *Chinese J. Cancer Res.* **30**, 669–676 (2018).

135. Li, Q. *et al.* Aldolase B Overexpression is Associated with Poor Prognosis and Promotes Tumor Progression by Epithelial-Mesenchymal Transition in Colorectal Adenocarcinoma. *Cell. Physiol. Biochem.* **42**, 397–406 (2017).

136. Qian, X. *et al.* Enolase 1 stimulates glycolysis to promote chemoresistance in gastric cancer. *Oncotarget* **8**, 47691–47708 (2017).

137. Esparza-Moltó, P. B. & Cuezva, J. M. The Role of mitochondrial H+-ATP synthase in cancer. *Frontiers in Oncology* **8**, (2018).

138. Olson, M. F. & Sahai, E. The actin cytoskeleton in cancer cell motility. *Clinical and Experimental Metastasis* **26**, 273–287 (2009).

139. Yamaguchi, H. & Condeelis, J. Regulation of the actin cytoskeleton in cancer cell

migration and invasion. *Biochimica et Biophysica Acta - Molecular Cell Research* **1773**, 642–652 (2007).

140. Lodish, H. F. *Molecular cell biology*. (W.H. Freeman, 2000).

141. Parker, A. L., Teo, W. S., McCarroll, J. A. & Kavallaris, M. An emerging role for tubulin isotypes in modulating cancer biology and chemotherapy resistance. *International Journal of Molecular Sciences* **18**, (2017).

142. Parker, A. L., Kavallaris, M. & McCarroll, J. A. Microtubules and their role in cellular stress in cancer. *Frontiers in Oncology* **4 JUN**, (2014).

143. Schmidt, M., Micke, P., Gehrmann, M. & Hengstler, J. G. Immunoglobulin kappa chain as an immunologic biomarker of prognosis and chemotherapy response in solid tumors. *Oncoimmunology* **1**, 1156–1158 (2012).

144. Chen, Z., Qiu, X. & Gu, J. Immunoglobulin expression in non-lymphoid lineage and neoplastic cells. *American Journal of Pathology* **174**, 1139–1148 (2009).

145. Ye, F., Chen, Y., Xia, L., Lian, J. & Yang, S. Aldolase A overexpression is associated with poor prognosis and promotes tumor progression by the epithelial-mesenchymal transition in colon cancer. *Biochem. Biophys. Res. Commun.* **497**, 639–645 (2018).

146. Chang, Y.-C., Yang, Y.-C., Tien, C.-P., Yang, C.-J. & Hsiao, M. Roles of Aldolase Family Genes in Human Cancers and Diseases. *Trends Endocrinol. Metab.* **29**, 549–559 (2018).

147. Chang, Y. C., Yang, Y. C., Tien, C. P., Yang, C. J. & Hsiao, M. Roles of Aldolase Family Genes in Human Cancers and Diseases. *Trends in Endocrinology and Metabolism* **29**, 549–559 (2018).

148. Azimzadeh, O., Atkinson, M. J. & Tapio, S. Qualitative and Quantitative Proteomic Analysis of Formalin-Fixed Paraffin-Embedded (FFPE) Tissue. in *Methods in molecular biology (Clifton, N.J.)* **1295**, 109–115 (2015).

8. Appendices

8.1 Appendix A: Optimization experiments of MALDI-IMS experiments

Tissue sectioning

After tissue sectioning the slides were placed on a hearing block for 10, minutes at 60°C. the slides were dewaxed with 3 serial washes of xylene and graded alcohol. The slides were finally washed with tap water to remove any access alcohol.

Slide coating protocol

- Mix 1:1 poly-L-lysine with water in an Eppendorf tube
- Add 1 µl of IGEPAL per 1.5 ml
- Streak out 2 drops of a Pasteur pipette onto an ITO coated slide using a tongue depressor
- Dry slide for 15 min on 60°C

Antigen retrieval

Tris -EDTA buffer solution was prepared with 2.42g Tris Base (Sigma, BCBS3963V) and 0.74g EDTA (Sigma, SLBH4839V) in 2 l of distilled water. The slides were pressure cooked for 1 min 30 seconds and air-dried overnight prior to processing.

Table 3. COMPOSITION OF SOLUTIONS

Haematoxylin 1g
Potassium/ammonium aluminium 50g
Sodium iodate 0,2g
Citric acid 1g
Chloral hydrate 50g
Distilled water 1000 ml

Table 4. REAGENT SUPPLIERS

Name	Supplier
Trifluoroacetic acid	Sigma
Acetonitrile	Merck
Water, proteomics grade	Fluka
HCCA matrix	Bruker
Absolute Ethanol	Millipore
Trypsin	Promeg
Acetonitrile	Sigma
ITO slides	Sigma
Poly-L-Lysine	Sigma
Ammonium Bicarbonate	Sigma-Aldrich
ITO coated slides	Sigma-Aldrich
Xylene	merck
Tris	Sigma
EDTA	Merck
Tween	Sigma

Matrix: Sinnapinic acid

The sinnapinc acid (SA) matrix (0.05g/l in 70% acetonitrile, 0.1 trifluoroacetic acid) was automatically spotted onto the tissue section using the SA spraying method on the HTX control software on the laptop which was connected to the HTX TM-sprayer. Teaching points were marked around the tissue section of interest using a tippex. A small section of the matrix was wiped off from the slide with 70% ethanol which was used to put the calibration spot. In an Eppendorf tube, a calibration standard mixture containing 6 µm/l protein standard was prepared (1.5 µm/l insulin (bovin), 10 µm/l apomyoglobin (equine), 10 µm cytochrome c (equine) with 6 µm/l SA matrix. Then 1µl of the calibration standard mixture was used create a spot on the slide area cleaned as mentioned above. The calibration spot was left to air dry at room temperature.

Teaching points were created on the slide by drawing an X mark in the region around the tissue.

- Draw X marks on the corners of the tissue slide making sure the tissue is within the marks (this will be used as teaching points downstream)
- Create a clean area on the slide by wiping off the matrix with 100% methanol for addition of the calibration spot
- Pick 1uL of the standards and matrix mixture and make a spot on the tissue. Allow to dry and make at least 4 layers. Then allow to dry completely.
- Position the slide into the slide holder and insert into the scanner
- Open CyberviewX application
- Scan settings should be at 3200dpi and 8-bit colour
- Open scan>prescan>prescan current frame
- Adjust the view area on the prescanned image
- Then click scan>scan current frame
- Open the new image that is automatically saved in D:data\imageScanner open the image with windows picture viewer
- Adjust the orientation according to the way the tissue slide will be loaded into the rapiflex

MALDI MSI

- Insert slide into Rapilflex slide holder
- Use a 96well plate lid to provide the orientation within the flex control environment
- Mark the tissue orientation where the X marks locations are on the 96 well plate lid
- Also highlight the location where the calibration spot is
- Place the cover lid on top of MTP384 ground steel plate and use it for orientation
- Open the flex control
- Insert the slide carrier into rapiflex
- Select the correct geometry (MTP slide adaptor II)
- Select method (LP_2-20kDa.par)
- Press the insert button
- Navigate to the calibration spot and press start to shoot.
- If the spectra is quality “add” repeat 3 times

Display sum spectra

- At the calibration tab select the relevant mass control list of calibrations. For proteins select “protein1Calib standard”
- Click on automatic assign and click on apply when at least 3 mass standards are assigned
- Open flex imaging
- Set up new imaging run
- Assign a name to your image run
- Acquisition settings raster 200um, method LP_2-20kDa.par
- Processing options select “perform smoothing” and perform base line subtraction
- Select the image that was scanned
- Create 3 teaching points on image at the fleximaging guided by the X marks that had been made
- On the flex imaging create ROIs either add polygon or rectangle around the areas where you want to analyse
- Draw the ROIs with the mouse then click “start automatic run” navigate to flex control and wait until the image run is complete
- Navigate back to flex imaging and click load results. The data will be saved in D:\data drive
- Inspect the loaded spectra

MALDI data analysis with ScilsLab

- Plug in scilslab software dongle
- Open the SCiLs Lab from the icon
- To make independent data sets Load the dataset into the SCiLs Lab by clicking on File>New>select TOF> Next
- Use the red+ to add the data sets you need to analyze
- To combine two or more data sets into one .sl file click on New (combine data sets into a new .sl file
- Select TOF as the instrument type> next>click on the red+ to navigate to the data sets stored in D:data file keeps adding as many as the data sets you want to compare
- When all the data sets have been imported then you can conduct downstream analysis
- Name the new data set formed based on the datasets merged

To create box plots and correlation plots

- Select the region of interested M/Z range and create an M/Z image
- When the M/Z image have been created click on the intensity box plot tab
- View the plots based on the selected region and the m/z interval image created
- Use this to compare different experiments based on the image runs that were loaded into the SCiLs Lab

Discovery of M/Z markers annotated on the tissue

Follow the online SCiLs Lab resource to conduct analysis of the imaging data [https://scils.de/mediacenter/]

8.2 Appendix B: LC-MS/MS experiments

2.1 Materials

Table 5. Solutions and Reagents

Xylene
Absolute ethanol
Lysis solution (LS): 0.1 M Tris-HCl, pH 8.0, 0.1 M DTT
20% (w/v) solution of SDS in water
Urea (UA) solution: 8 M urea (Sigma, U5128) in 0.1 M Tris/HCl pH 8.5
IAA solution: 0.05 M iodoacetamide in UA
Trypsin, stock 0.4 µg/µL
Ammonium bicarbonate (ABC) solution: 0.05M NH4HCO3 in water

Note: LS, UA, and IAA solutions must be freshly prepared and used within a day.

Table 6. Equipment

Ultra-Turrax blender (T 10 basic Ultra, IKA, Staufen, Germany).
Branson Sonifier (Heinemann, Schwäbisch Gmünd, Germany)
Vivacon 500 (Sartorius Stedim Biotech)
Refrigerated bench-top centrifuge, set to 18°C
Incubator set to 37°C
Wet chamber with a rack for Eppendorf tubes
Heating block with agitation, set to 99°C
Centrifugal vacuum-dryer (Speed-Vac)

2. 2 FASP Protocol

- Incubate FFPE tissues slices in 1 mL of xylene in an Eppendorf-type tube with gentle agitation at room temperature for 5 min.
- Remove the solution, add 1 mL of xylene and incubate as in (1).
- Remove the solution and repeat steps 1 and 2 in each using 1 mL of absolute ethanol.
- Remove ethanol and vacuum-dry the sample.
- Mix the dried tissue with LS in a tissue to buffer ratio of 1:20 and homogenize on ice using the disperser for 3 min.
- Sonicate the suspension on ice for 3 min. (output control 5; duty cycle 20%)
- Add 20% SDS to the suspension to a final concentration of 4%.
- Incubate in a heating block with agitation (600 rpm) at 99°C for 60 min.
- Remove the tube from the heating block and let it slowly chill to the room temperature.
- Centrifuge the extract at 16,000 × g for 10 min.
- mix up to 50 µL of the clarified lysate with 200 µL of UA in the filter unit and centrifuge at 14,000 × g for 30 min.
- Add 200 µL of UA to the filter unit and centrifuge at 14,000 × g for 20 min.
- Discard the flow-through form the collection tube.

- Add 100 µL IAA solution and mix at 600 rpm in a thermo-mixer for 1 min and incubate without mixing for 20 min.
- Centrifuge the filter units at 14,000 × g for 10 min.
- Add 100 µL of UA to the filter unit and centrifuge at 14,000 × g for 15 min. Repeat this step twice.
- Add 100 µL of ABC to the filter unit and centrifuge at 14,000 × g for 10 min. Repeat this step twice.
- Add 40 µL ABC with trypsin (enzyme to protein ratio 1:100) or another protease and mix at 600 rpm in thermo-mixer for 1 min.
- Incubate the units in a wet chamber at 37°C for 4 -18 h.
- Transfer the filter units to new collection tubes.
- Centrifuge the filter units at 14,000 × g for 10 min.
- Add 50 µL of ABC and centrifuge the filter units at 14,000 × g for 10 min.
- (optional) Measure the peptide yield in the filtrate using UV spectrophotometer.

8.3 Appendix C.1: Region of interest (ROIs) drawn on H&E slides and scanned with MALDI-IMS

8.3.1 Region of interest (ROIs) drawn on H&E slides and scanned with MALDI-IMS

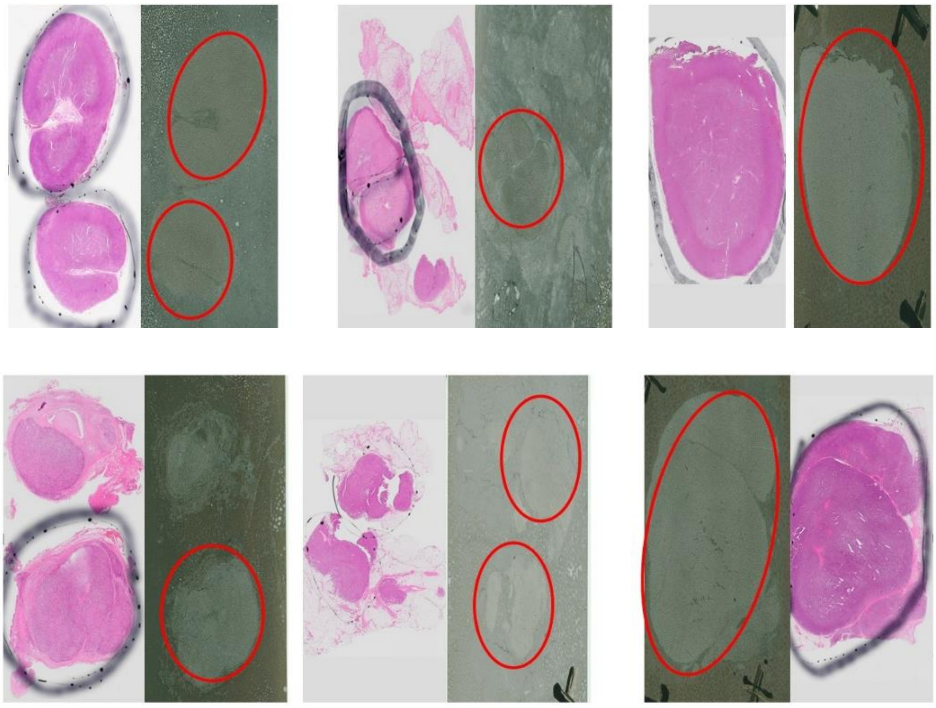


Figure 17 ROIs drawn on ABC HIV Negative and GCB HIV negative tissue slides.

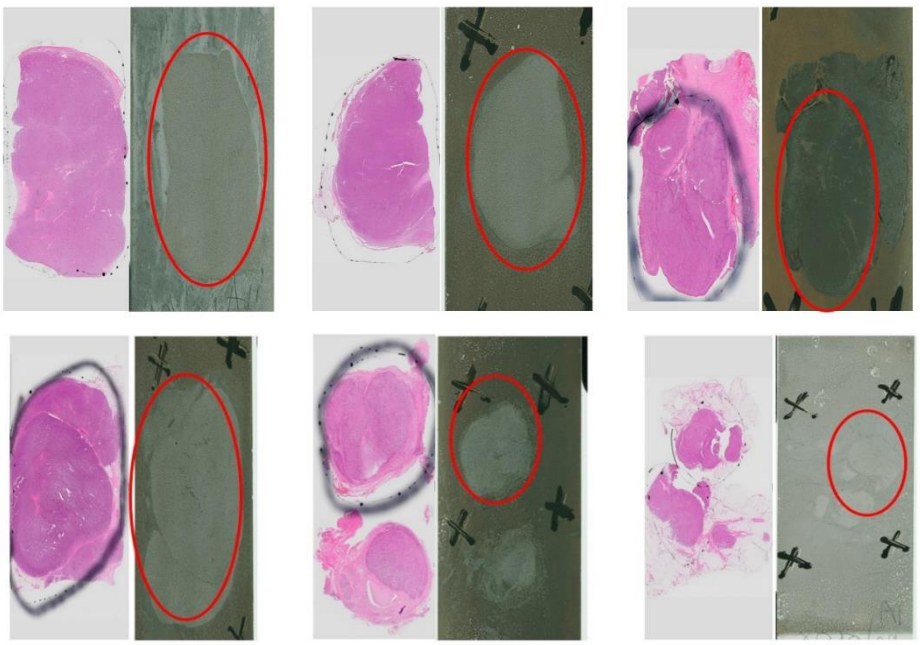


Figure 18. ABC HIV positive and GCB HIV positive

8.3. Appendix C.2: Spatial distribution of the ion with the highest intensity

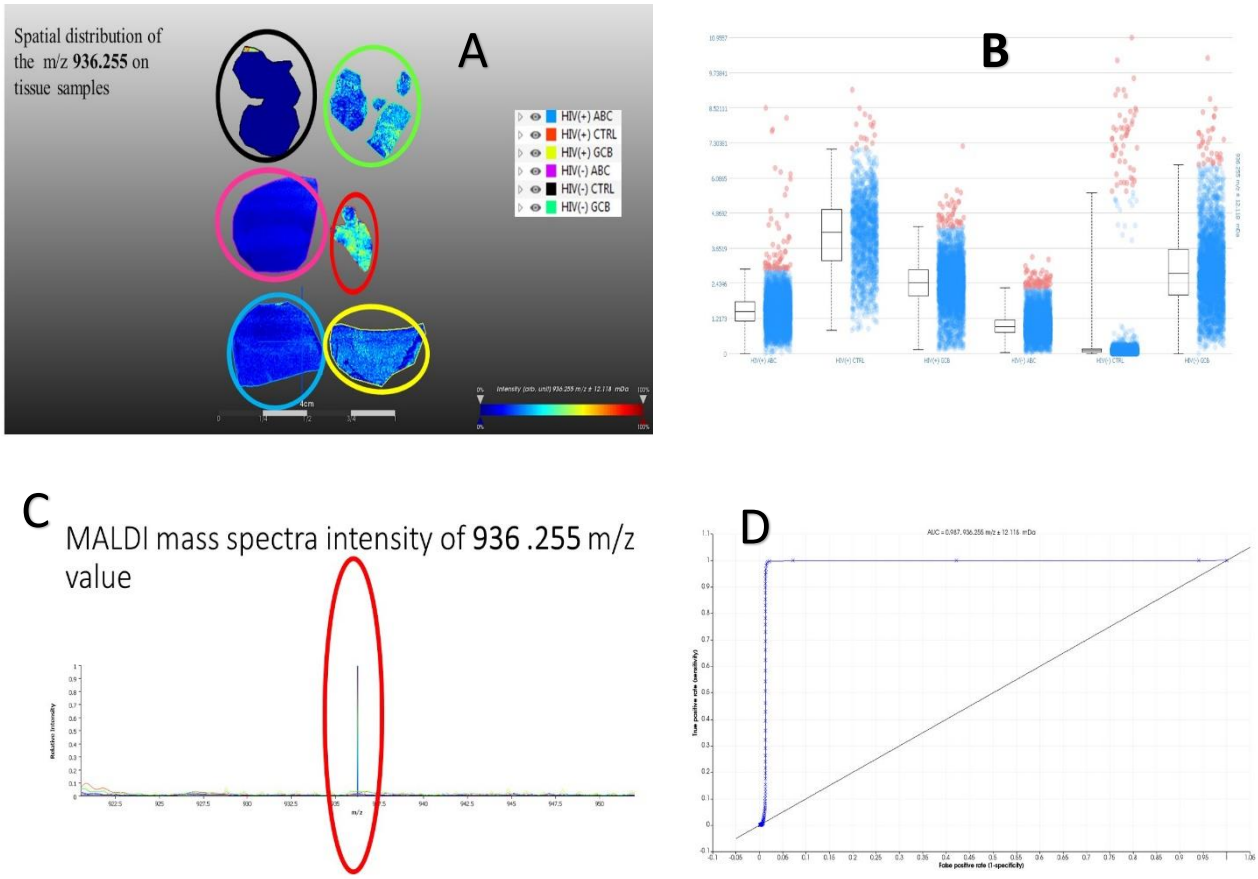


Figure 19. Spatial distribution of the selected m/z 0.967. (A) m/z 936.255 had the highly distributed on the HIV (-) control (indicated in red). (B) intensity box plot of the selected m/x values. (C) Average peak MALDI mass spectra intensity of all the samples. (D) Receiver Operating characteristic (ROC) curve analysis of selected ion (AUC=0,967).

8.3.3 Appendix C.3 ROIS DRAWN IN MALDI IMS ANALYSIS

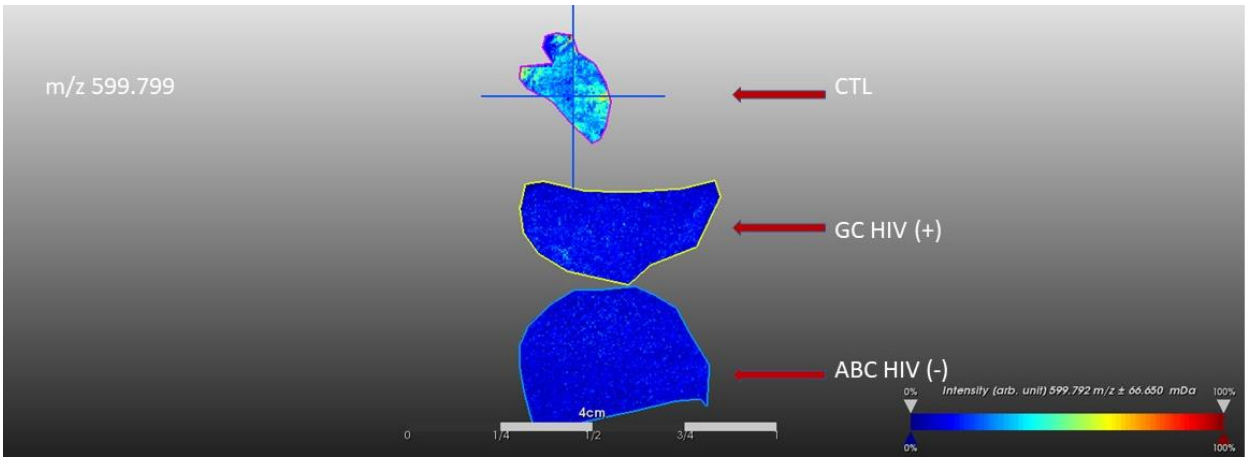


Figure 20. Spatial distribution of ions discriminating Control tissue (CTL), and Sample 1; GC HIV (+), Sample 2; ABC HIV (-) of the MALDI-imaging data. The m/z-value 599.799 ± 0,067% (AUC=0.973) is abundant in the CTL tissue. Visualization was performed in SCiLS

8.4 Appendix D: Exclusive ion mass (m/z) values identified in DLBCL cases

Table 7 The number of different ion mass (m/z) values identified in DLBCL cases.

HIV positive DLBCL GCB Ion mass(m/z)	**HIV negative DLBCL GCB Ion mass(m/z)**	**HIV positive DLBCL ABC Ion mass(m/z)**	**HIV negative DLBCL ABC Ion mass(m/z)**
2086.648	3025.453	1030.84	2042.259
2042.731	3025.332	884.607	1837.372
1837.687	686.579	860.685	860.115
1837.651	665.747	859.642	649.315
1837.444	649.617	729.781	649.29
1065.886	649.496	665.759	649.266
1059.936	643.716	643.679	649.242
1059.9	643.692	627.683	648.903
860.636	643.655	1030.84	643.316

854.759		884.607	2042.259
854.662		860.685	1837.372
854.529		854.323	860.115
854.444		838.774	649.315
838.944		729.781	649.29
838.871		665.759	649.266
665.905		664.705	649.242
665.541		643.679	648.903
649.92		627.683	643.316
649.533			
649.484			
649.424			

8.5 Appendix E: List of differentially expressed proteins

Table 8. Identification information of differentially expressed proteins

No. of Peptides	Protein ID	Protein name	Gene	pathway
7	G3V1U9	Tubulin Alpha	TUBA1A	cell division
5	A8K092	ATP synthase	ATP5F1A	ATP synthesis
6	P06733	Enolase	ENO1	GLYCOLYSIS
13	P60709	Actin	ACTB	ATP binding
11	P14618	Pyruvate Kinase	PKM	GLYCOLYSIS
5	P01834	Ig kappa chain	IGKC	B-cell signalling receptor

8.6 Appendix F: Exclusive protein ID's for ABCN

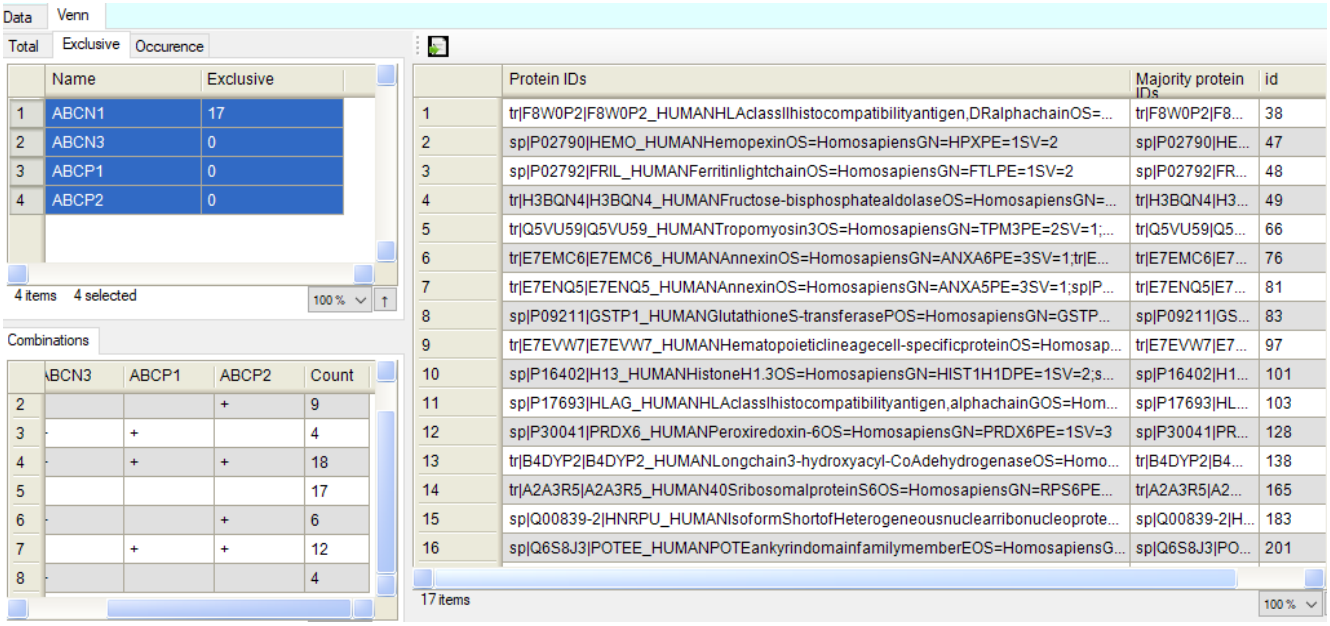


Figure 21 Numeric Venn Diagram ABC subtype

8.7 Appendix G: Two-sample t-test Statistical analysis for GCBN and GCBP.

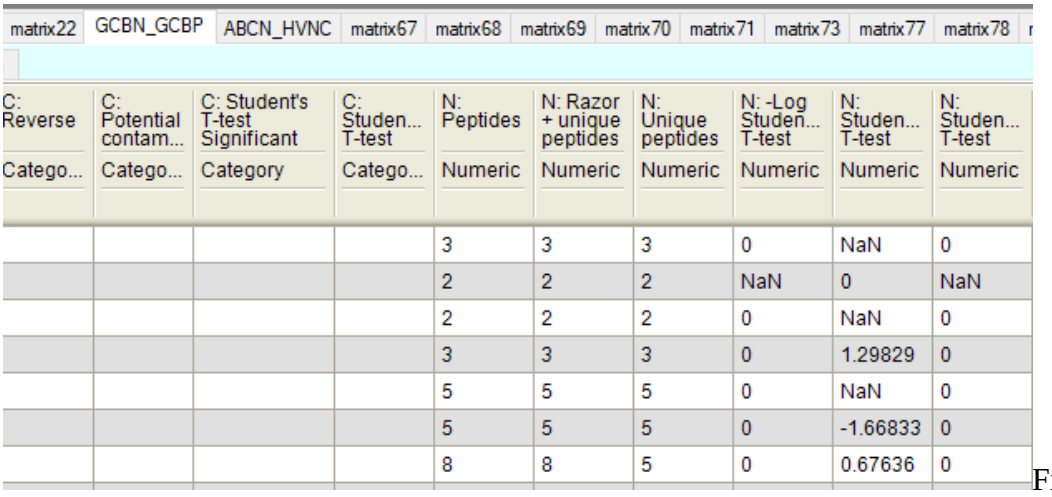


Figure 22 Two-sample t-test Statistical analysis for GCBN and GCBP. No protein was significantly different between HIV negative DLBCL GCB subtypes and HIV positive DLBCL GCB subtype.

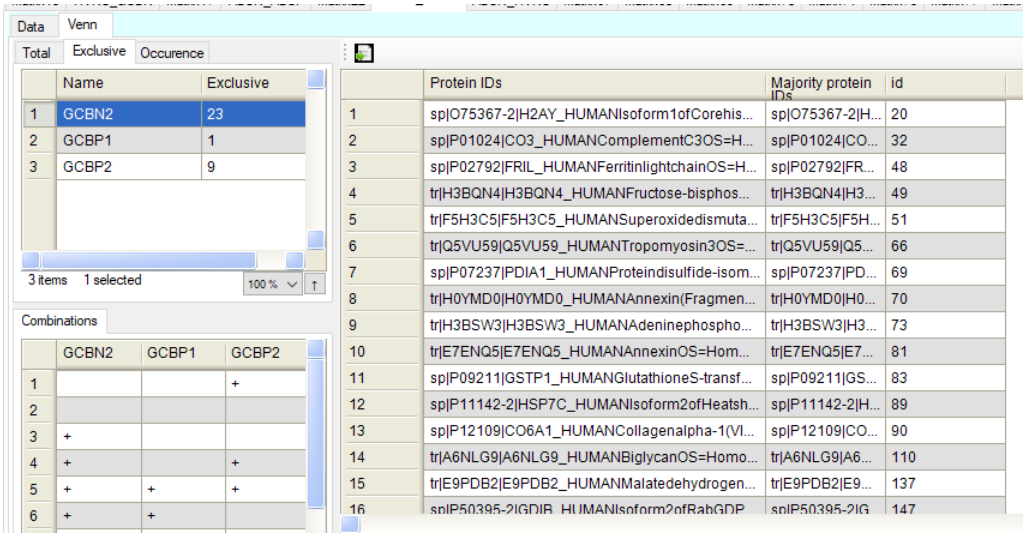


Figure 23 Numeric Venn diagram showing exclusive proteins. In GCBN and GCBP subtypes

A Student t test HIV negative control (HVNC and ABCN) and (HVNC and GCBN) subtypes, and no protein was found to significantly different (Appendix G, figure.21 and 22).

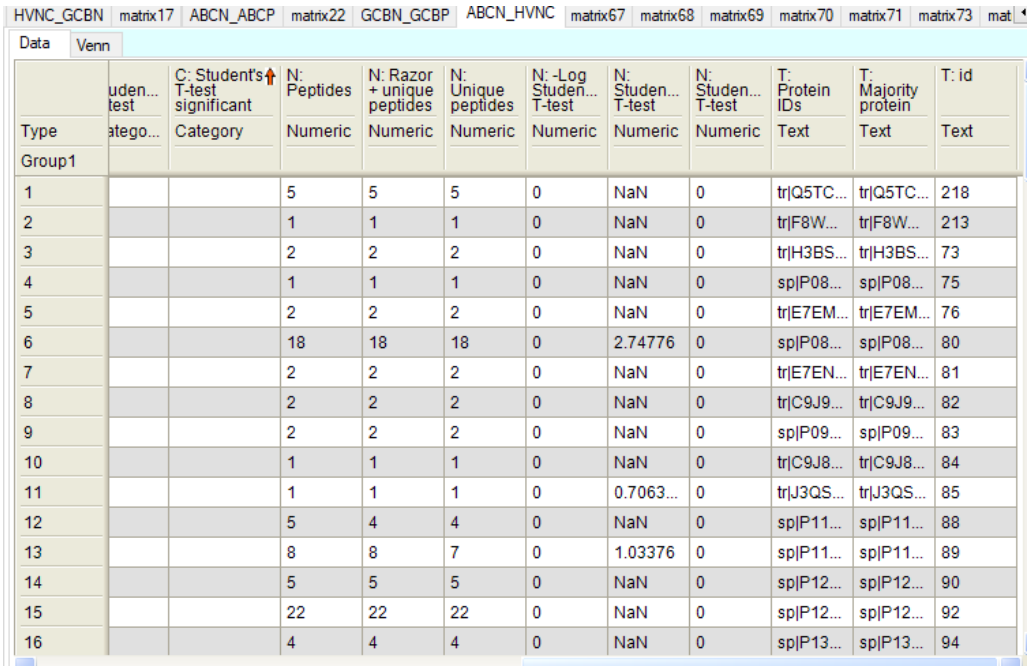
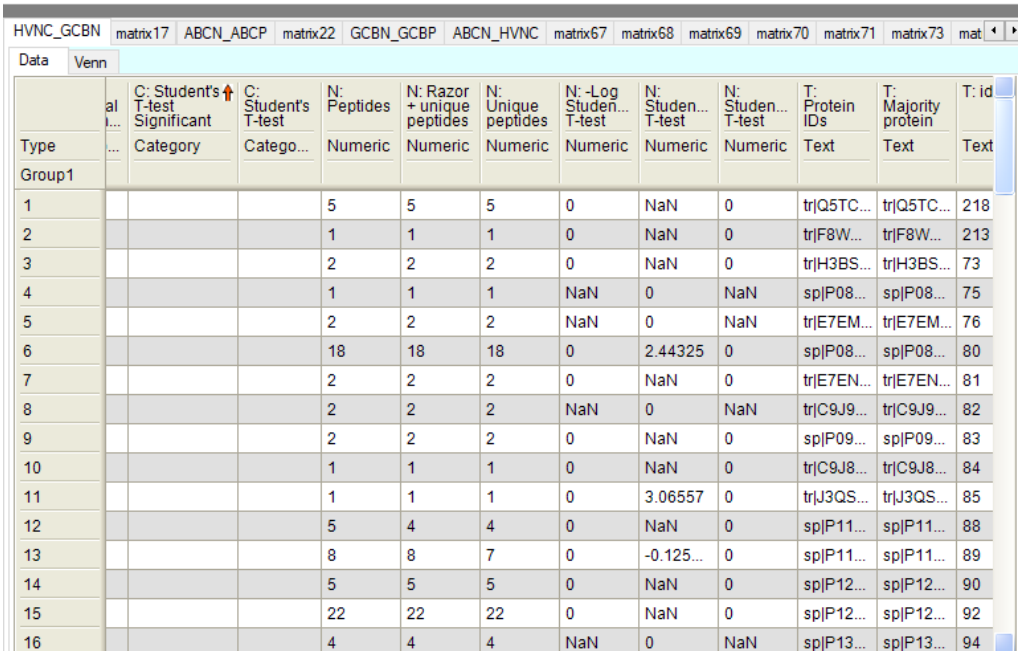


Figure 24 Two-sample t-test Statistical analysis for HVNC and ABCN. No protein was significantly different between HIV negative control HIV negative DLBCL ABC subtypes.



HVNC_GCBN | matrix17 | ABCN_ABCP | matrix22 | GCBN_GCBP | ABCN_HVNC | matrix67 | matrix68 | matrix69 | matrix70 | matrix71 | matrix73 | mat

Data | Venn

	al ...	C: Student's T-test Significant	C: Student's T-test	N: Peptides	N: Razor + unique peptides	N: Unique peptides	N: -Log Studen... T-test	N: Studen... T-test	N: Studen... T-test	T: Protein IDs	T: Majority protein	T: id
Type	...	Category	Catego...	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Text	Text	Text
Group1												
1				5	5	5	0	NaN	0	tr|Q5TC...	tr|Q5TC...	218
2				1	1	1	0	NaN	0	tr|F8W...	tr|F8W...	213
3				2	2	2	0	NaN	0	tr|H3BS...	tr|H3BS...	73
4				1	1	1	NaN	0	NaN	sp|P08...	sp|P08...	75
5				2	2	2	NaN	0	NaN	tr|E7EM...	tr|E7EM...	76
6				18	18	18	0	2.44325	0	sp|P08...	sp|P08...	80
7				2	2	2	0	NaN	0	tr|E7EN...	tr|E7EN...	81
8				2	2	2	NaN	0	NaN	tr|C9J9...	tr|C9J9...	82
9				2	2	2	0	NaN	0	sp|P09...	sp|P09...	83
10				1	1	1	0	NaN	0	tr|C9J8...	tr|C9J8...	84
11				1	1	1	0	3.06557	0	tr|J3QS...	tr|J3QS...	85
12				5	4	4	0	NaN	0	sp|P11...	sp|P11...	88
13				8	8	7	0	-0.125...	0	sp|P11...	sp|P11...	89
14				5	5	5	0	NaN	0	sp|P12...	sp|P12...	90
15				22	22	22	0	NaN	0	sp|P12...	sp|P12...	92
16				4	4	4	NaN	0	NaN	sp|P13...	sp|P13...	94

Figure 25. Two-sample t-test Statistical analysis for HVNC and ABCN. No protein was significantly different between HIV negative control HIV negative DLBCL GCB subtypes.

8.7 Appendix H: Perseus data processing pipeline

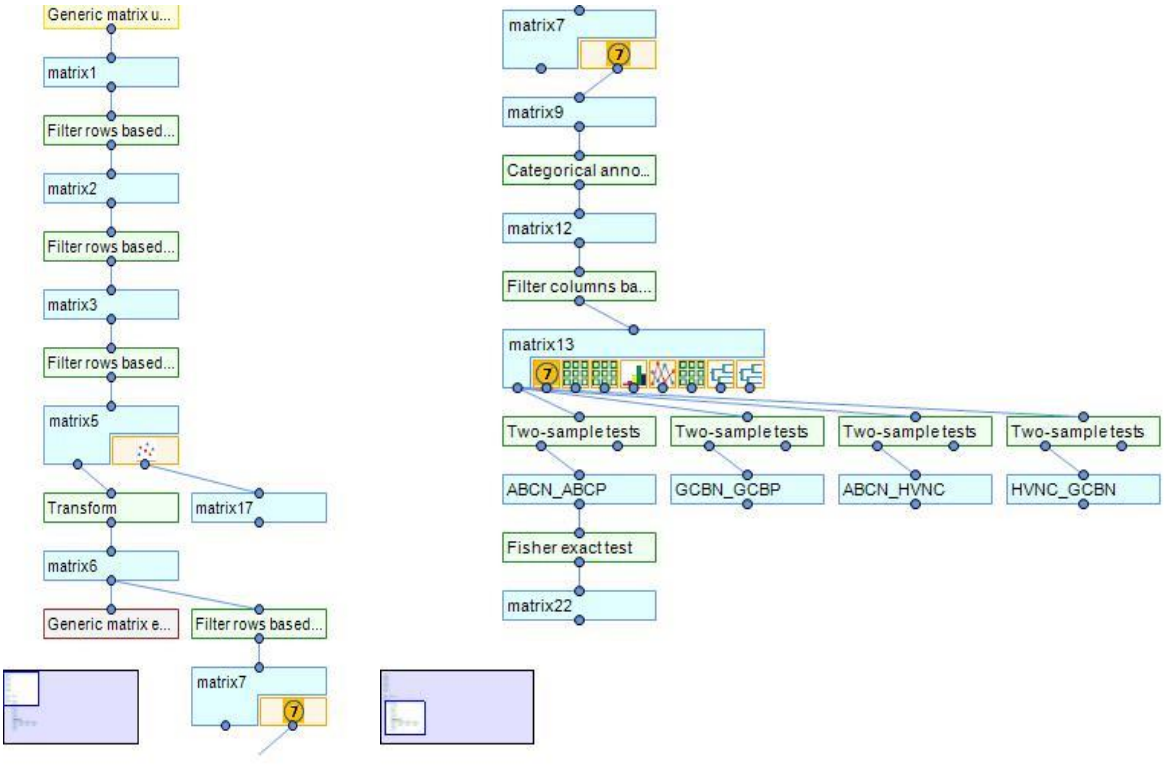


Figure 26. Perseus data processing pipeline

8.8 Appendix I. A list of extranodal sites

Table 9. A Breakdown of extranodal sites found in the HIV positive and HIV negative cases.

HIV status	Subtype	Site	Specificity
NEGATIVE	ABC	extranodal	submandibular
NEGATIVE	ABC	extranodal	lymph node
NEGATIVE	ABC	extranodal	axilla
NEGATIVE	GCB	extranodal	mensenteric
NEGATIVE	GCB	extranodal	supraclavicular
NEGATIVE	GCB	extranodal	axillary
POSITIVE	ABC	extranodal	testis
POSITIVE	ABC	extranodal	axillary
POSITIVE	ABC	extranodal	inguinal
POSITIVE	GCB	extranodal	mesenteric
POSITIVE	GCB	extranodal	right testis
POSITIVE	GCB	extranodal	axillary

8.9 Appendix J. A list of potential protein biomarkers

Table 10 : List of Potential Biomarkers

Potential biomarker ID	Proteins name
Q99880	Histone
D6RD66	Wdrepeat
O75367-2	Isoform1 of Core histone macro
B7Z7A9	Phosphoglycerate kinase
P01024	Complement-C3
P01834	Ig kappa chain-C region
P01857	Ig gamma-4chain-C region
F8W0P2	HLA class-II histocompatibility antigen
D6RAQ3	Prelamin-A
P02647	ApolipoproteinA
P02790	Hemopexin
P02792	Ferritin light chain
P04075	Fructose-bisphosphate aldolase
F5H3C5	Super oxided ismutase[Mn]
E7EUT4	Glyceraldehyde-3-phosphatedehydrogenase
P04792	Heatshockproteinbeta-1

Q99878	HistoneH2Atype1
P06576	ATPsynthasesubunitbeta
P06733	Alpha-enolase
Q5VU59	Tropomyosin3
P07237	Protein disulfide-isomerase
H0YMD0	Annexin(Fragment)
F8VUJ7	Tubulinbetachain
H3BSW3	Adenine phosphoribosyl transferase
P08123	Collagenalpha-2(I)chain
E7EMC6	Annexin
P08670	Vimentin
E7ENQ5	Annexin
C9J9K3	40S ribosomal proteinSA
P09211	GlutathioneS-transferaseP
C9J8F3	Fructose-bisphosphatealdolaseC
J3QSA3	Polyubiquitin-B
P11021	78kDa glucose-regulated protein
P11142-2	Isoform2 of Heat shockcognate 71kDa protein
P12109	Collagen alpha-1(VI)chain
P12111-2	Collagen alpha-3(VI)chain
P13639	Elongation factor2
P13796	Plastin-2
E7EVW7	Hematopoietic lineage cell-specific protein
P14618	Pyruvate kinase isozymes M1/M2
P16401	HistoneH1.5
P16402	Histone H1.3
P17693	HLA classI histocompatibility antigen, alpha chain G
C9JGI3	Thymidinephosphorylase
E9PBF6	Lamin-B1
Q5HY54	Filamin-A
A6NLG9	Biglycan
P22626-2	IsoformA2 of Heterogeneous nuclearribonucleo proteinsA2/B1
P23528	Cofilin-1
P24821-5	Isoform 5 of Tenascin
A8K092	ATP synthases ubunitalpha, mitochondrial
P26038	Moesin
P30041	Peroxiredoxin-6
P30086	Phosphatidylethanolamine-bindingprotein1
P31146	Coronin-1A

F5GZT4	Heterogeneous nuclear ribonucleo proteinH
P35579	Myosin-9
E9PDB2	Malatede hydrogenase,mitochondrial
B4DYP2	Longchain3-hydroxyacyl-CoAdehydrogenase
P42224-2	Isoform Beta of Signal transducer andactivator of transcription1-alpha/beta
F5H571	Ubiquitincarboxyl-terminalhydrolase5(Fragment)
P50395-2	Isoform2 of Rab GDP dissociation inhibitor beta
P52272-2	Isoform2 of Heterogeneous nuclear ribonucleo proteinM
P55072	Transitional endoplasmicreticulum ATPase
F8WDD7	Actin-relatedprotein2
P60174-1	Isoform2 of Triosephosphateisomerase
P60709	Actin,cytoplasmic1
P63267	Actin,gamma-entericsmoothmuscle
A2A3R5	ribosomal proteinS6
P62820-2	Isoform2 of Ras-related proteinRab-1A
P62826	GTP-binding nuclear protein Ran
P62937	Peptidyl-prolylcis-transisomeraseA
B0AZS6	14-3-3proteinzeta/delta
Q5VTE0	Putative elongation factor1-alpha-like3
G3V1U9	Tubulinalpha-1A chain
P68871	Hemoglobin sub unitbeta
P69905	Hemoglobinsubunitalpha
Q00839-2	IsoformShort of Heterogeneous nuclear ribonucleo proteinU
Q01518	Adenylylcyclase-associated protein1
H0YNE3	Proteasome activator complex subunit1
Q07666-3	Isoform3 of KH domain-containing,RNA-binding,signal transduction-associated protein1
Q13263-2	Isoform2 of Transcription intermediary factor1-beta
G8JLA8	Transforming growth factor-beta-induced protein ig-h3
Q6S8J3	POTE ankyrin domain family member E
Q86UX7-2	Isoform2 of Fermitin familyhomolog 3
Q86VP6-2	Isoform2 of Cullin-associated NEDD8-dissociated protein1
F8W914	Isoform5 of Reticulon-
Q5TCU6	Talin1